

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**EFEK PENGGABUNGAN *OUTLIER* DAN *UNCLASSIFIED*
DATA PADA PEMODELAN *ROUGH-REGRSI*
(Studi Kasus : Tiga Variabel Bebas Numerik)**

TUGAS AKHIR

Diajukan sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Sains pada
Program Studi Matematika

Oleh :

IRMA SURYANI
11654203678



UIN SUSKA RIAU

UIN SUSKA RIAU

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2019**

LEMBAR PERSETUJUAN

EFEK PENGGABUNGAN *OUTLIER* DAN *UNCLASSIFIED* DATA PADA PEMODELAN *ROUGH-REGRESI* (Studi Kasus : Tiga Variabel Bebas Numerik)

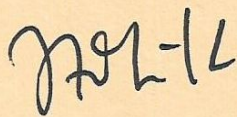
TUGAS AKHIR

Oleh:

IRMA SURYANI
11654203678

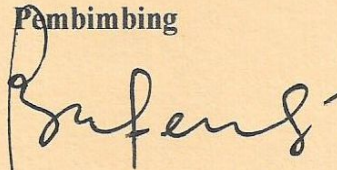
Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 17 Desember 2019

Ketua Program Studi



Ari Pani Desvina, M.Sc.
NIP. 19811225 200604 2 003

Pembimbing



Dr. Riswan Efendi, M.Sc.
NIP. 19781025 200604 1 001

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PENGESAHAN

EFEK PENGGABUNGAN *OUTLIER* DAN *UNCLASSIFIED* DATA PADA PEMODELAN *ROUGH-REGRESI* (Studi Kasus : Tiga Variabel Bebas Numerik)

TUGAS AKHIR

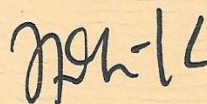
Oleh:

IRMA SURYANI
11654203678

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 17 Desember 2019

Pekanbaru, 17 Desember 2019
Mengesahkan

Ketua Program Studi



Ari Pani Desvina, M.Sc.
NIP. 19811225 200604 2 003



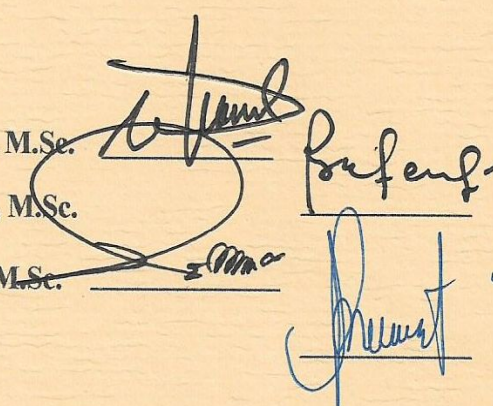
DEWAN PENGUJI

Ketua : Mohammad Soleh, M.Sc.

Sekretaris : Dr. Riswan Efendi, M.Sc.

Anggota I : Dr. Rado Yendra, M.Sc.

Anggota II : Rahmadeni, M.Si.



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau serta terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam tugas akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 17 Desember 2019

Yang membuat pernyataan,

IRMA SURYANI
11654203678

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR PERSEMBAHAN

*Tiada tempat yang pantas mengadu kecuali pada-Mu,
Tiada tempat yang layak untuk meminta kecuali pada-Mu,
Kini kuberSyukur ya Allah atas kelulusan yang kau berikan padaku.
Untuk Rasulullah Shallallahu „alaihi wasallam terima kasih atas tauladan mu*

« Motivasi Menuntut Ilmu »

“Harta Yang Tak Pernah Habis Adalah Ilmu Pengetahuan dan Ilmu Yang Tak Ternilai Adalah Pendidikan”

« Motivasi hidup »

Don't give up when you still have something to give. Nothing is really over until the moment you stop trying”

(Irma Suryani)

Kupersembahkan karya yang berharga ini untuk. . .

Ayahanda dan Ibunda tercinta dengan lautan kasih dan sayangnya Yang selalu tercurah lewat doa dan pengorbanan yang tulus, Yang telah mengajari aku ikhlas ketika air mata berbicara tentang kesedihan. Terima kasih telah mengajari aku menggunakan pena sehingga dapat ku tulis gemericik air, udara dingin, kabut senja sampai daun gugur. Setiap jerih payah dan bulir tetesan keringatmu akan menjadi saksi betapa berharganya pengorbananmu dan Terimakasih juga telah mendo'akanku di setiap sujudmu tanpa mengenal lelah.

Untuk Pembimbingku (Dr. Riswan Efendi, M.Sc.)

Terimakasih untuk segala rasa sabar, ikhlas dalam membimbing diri ini yang masih sering lalai dan banyak kekurangan. Terimakasih telah memberikan masukan dan motivasi untuk penyelesaian tugas akhir ini.

Untuk Semua dosen Program Studi Matematika FST;

Terimakasih untuk semua ilmu-ilmu yang diajarkan selama saya masih duduk di bangku kuliah dan nasehat serta motivasinya.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

EFEK PENGGABUNGAN *OUTLIER* DAN *UNCLASSIFIED* DATA PADA PEMODELAN *ROUGH-REGRESI* (Studi Kasus : Tiga Variabel Bebas Numerik)

IRMA SURYANI
11654203678

Tanggal Sidang : 17 Desember 2019
Periode Wisuda :

Program Studi Matematika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Pekanbaru

ABSTRAK

Model *rough-regresi* pada penelitian ini diaplikasikan pada data yang dibangkitkan melalui teknik simulasi dan diimplementasikan ke data real. Namun, sering kali dalam penerapan model regresi ditemukan satu atau beberapa titik data yang berada jauh dari pola data pada umumnya atau dikenal dengan *outlier*. Munculnya *outlier* dapat mempengaruhi *R-Square* yang diperoleh. *Outlier* pada model regresi sama halnya dengan *unclassified* data pada model *rough sets*. *R-Square outlier* yang diperoleh dari model regresi rata-rata lebih kecil dibandingkan dengan *R-Square* yang diperoleh dari *unclassified* data. Maka melalui penelitian ini penulis tertarik untuk melihat pengaruh dari penggabungan *outlier* dan *unclassified* data dengan memperhatikan *R-Square* dari setiap percobaan data simulasi. Hasil simulasi menunjukkan bahwa data simulasi dari penggabungan *outlier* dan *unclassified* data rata-rata memiliki *R-Square* yang meningkat. Sedangkan pada implementasi data real, penggabungan *outlier* dan *unclassified* data juga mengalami peningkatan nilai *R-Square*. Secara umum dapat disimpulkan bahwa hasil yang diperoleh melalui kedua prosedur ini adalah sejalan. Dalam hal ini penggabungan antara *outlier* dan *unclassified* data lebih mampu meningkatkan *R-Square*. Sehingga dapat diartikan bahwa model *rough-regresi* memiliki kontribusi yang baik dibandingkan dengan model regresi biasa.

Kata kunci: Data Simulasi, *Outlier*, *R-Square*, *Rough-regresi*, *Unclassified* Data

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

THE EFFECT OF MERGING OUTLIERS AND UNCLASSIFIED DATA ON ROUGH-REGRESSION MODELING

(Case Study : Three Numerical Free Variables)

IRMA SURYANI

11654203678

Date of Final Exam : 17 December 2019

Date of Graduation :

*Department of Mathematics
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Jl. HR. Soebrantas No. 155 Pekanbaru*

ABSTRACT

The rough-regression model in this study was applied to data generated through simulation techniques and implemented to real data. However, often in applying the regression model one or several data points are found that are far from the data patterns in general or known as outliers. The emergence of outliers can affect the R-Square obtained. Outliers in the regression model are the same as unclassified data in the rough sets model. R-Square outliers obtained from the average regression model are smaller than R-Square obtained from unclassified data. So through this study the authors are interested in seeing the effect of merging outliers and unclassified data by paying attention to the R-Square of each simulated data experiment. Simulation results show that the simulation data from the merging of outliers and unclassified data on average has an increased R-Square. Whereas in the implementation of real data, the merging of outliers and unclassified data also increased the value of R-Square. In general it can be concluded that the results obtained through these two procedures are in line. In this case the merging between outliers and unclassified data is better able to increase R-Square. So it can be interpreted that the rough-regression model has a good contribution compared to the ordinary regression model.

Keywords: *Simulation Data, Outlier, R-Square, Rough-regression, Unclassified Data*

KATA PENGANTAR

Alhamdulillahirabbil'alamin

Segala puji Allah *subhanahu waata'ala* atas yang senantiasa melimpahkan rahmat, karunia, dan petunjuk-Nya lah penulis bisa menyelesaikan tugas akhir ini. Shalawat beriringan salam kepada Nabi Muhammad *shallallahu 'alaihi wasallam* yang telah membawa kita dari zaman yang tidak berpengetahuan sampai zaman yang memiliki kemajuan ilmu dan teknologi yang kita rasakan pada saat ini.

Penelitian ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar sarjana sains dan teknologi pada program studi matematika. Dalam penyusunan dan penyelesaian penelitian ini, penulis banyak sekali mendapat bimbingan, bantuan, arahan, nasehat, petunjuk, perhatian serta semangat dari berbagai pihak terutama orang tua tercinta yang tidak pernah lelah dan tiada henti melimpahkan kasih sayang, perhatian, motivasi yang membuat penulis mampu untuk terus dan terus melangkah, pelajaran hidup, juga materi yang tak mungkin bisa terbalas. Jasa-jasamu kan selalu kukenang hingga akhir hayatku dan semoga Allah menjadikan jasa-jasamu sebagai amalan soleh, Aamiin. Kemudian penulis juga mengucapkan terima kasih yang tak terhingga kepada:

1. Bapak Prof. Dr. H. Akhmad Mujahidin, M.Ag., selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Drs. Ahmad Darmawi, M.Ag., selaku Dekan Fakultas Sains dan Teknologi.
3. Ibu Ari Pani Desvina, M.Sc., selaku Ketua Program Studi Matematika dan Pembimbing Akademik.
4. Ibu Fitri Aryani, M.Sc., selaku Sekretaris Program Studi Matematika.
5. Bapak Dr. Riswan Efendi, M.Sc., selaku Pembimbing Tugas Akhir yang telah memberi bimbingan, pengarahan serta ilmunya.
6. Bapak Dr. Rado Yendra, M.Sc dan ibu Rahmadeni, M.Si, selaku Penguji yang telah memberikan kritikan dan saran sehingga selesainya tugas akhir ini.

Hak Cipta Dilindungi Undang-Undang

7. Seluruh Dosen Matematika, Fakultas Sains dan Teknologi yang telah banyak memberi nasehat, bimbingan, serta bantuan kepada penulis.
8. Keluarga tercinta, yang telah memberikan motivasi, dukungan, do'a, dan materi yang tak henti-hentinya serta kasih sayang yang sangat tulus kepada penulis.
9. Saudara-saudara Penulis (Kakak Aryati Citra, Adik Muhammad Rafi, dan Adik Amelia Putri) yang selalu memberikan dukungan dan kasih sayang kepada penulis.
10. Teman-teman TA Squad (Anisa Rahmah B, Siska Khairunnisa, Yuli Wahyuni Zelvy, Intan Eria Elfi, Mutiatul Hasanah dan Reza Chairisman) yang selalu memberi dukungan dan motivasi kepada penulis.
11. Seluruh teman-teman seperjuangan Program Studi Matematika angkatan 2016 terkhusus lokal B.
12. Teman-teman KKN Desa Gunung Mulya, Kabupaten Kampar, Kecamatan Gunung Sahilan.
13. Semua pihak yang telah banyak membantu baik secara langsung maupun tidak langsung dalam penyelesaian penelitian ini yang tidak dapat penulis sebutkan satu persatu.

Semoga kebaikan yang telah mereka berikan kepada penulis menjadi amal kebaikan dan mendapat balasan yang setimpal dari Allah *subhanahu waata'ala*. Aamiin. Dalam penulisan ini penulis menyadari bahwa penelitian tugas akhir ini belum sempurna. Namun, penulis sudah berusaha untuk mencapai hasil yang maksimal. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan tugas akhir ini. Akhir kata penulis harap semoga penelitian tugas akhir ini dapat bermanfaat bagi penulis dan pihak-pihak yang memerlukan.

Pekanbaru, 17 Desember 2019

Irma Suryani

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbarik sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

	Halaman
LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xvi
 BAB I PENDAHULUAN	
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-3
1.3 Batasan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-3
1.6 Sistematika Penulisan	I-4
 BAB II LANDASAN TEORI	
2.1 Konsep Regresi Linier	II-1
2.1.1 Bentuk Umum Model Regresi Linier Berganda	II-2
2.1.2 Asumsi-asumsi Model Regresi Linier Berganda	II-4
2.1.3 Pengujian Parameter Regresi Linier Berganda	II-7
2.2 Konsep <i>Rough Sets</i>	II-8
2.3 Konsep <i>Data Outlier</i>	II-12
2.4 Konsep <i>Unclassified Data</i>	II-16

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.5	Konsep Data <i>Training</i> dan <i>Testing</i>	II-16
2.6	Konsep Metode Simulasi	II-16
2.7	Kajian Penelitian Sebelumnya	II-16

BAB III METODOLOGI PENELITIAN

BAB IV HASIL DAN PEMBAHASAN

4.1	Data Simulasi Tingkat Kemiskinan.....	IV-1
4.2	Implementasi Data Tingkat Kemiskinan.....	IV-17

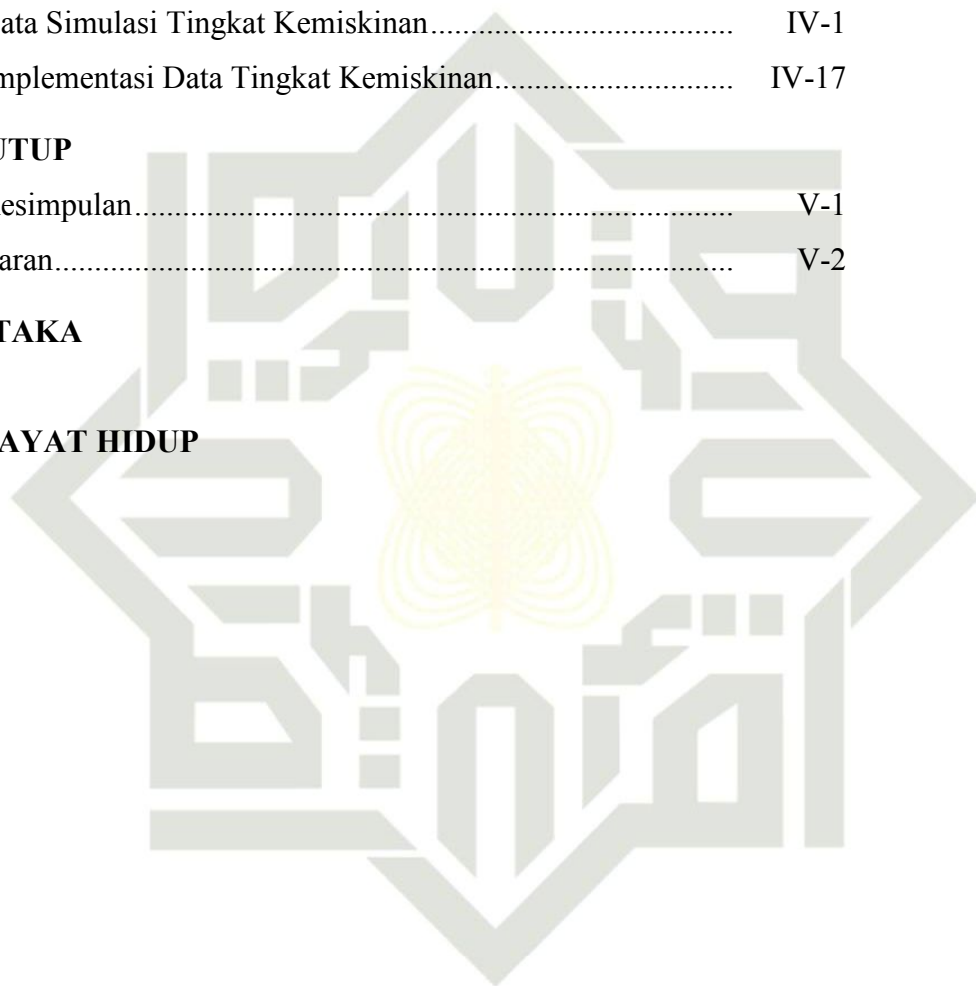
BAB V PENUTUP

5.1	Kesimpulan.....	V-1
5.2	Saran.....	V-2

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR RIWAYAT HIDUP



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1 <i>Scatter-Plot</i> dari Data dengan Observasi ke- i	II-14
Gambar 2.2 Skema Identifikasi Pencilan Menggunakan IQR atau <i>Boxplot</i>	II-15
Gambar 3.1 <i>Error</i> yang Dibangkitkan	III-2
Gambar 3.2 Prosedur untuk <i>Genereted Data</i>	III-3
Gambar 3.3 <i>Flowchart</i> Implementasi Model <i>Rough-Regresi</i>	III-5
Gambar 4.1 Membangkitkan <i>Error</i>	IV-2
Gambar 4.2 Membangkitkan X_1, X_2, X_3	IV-2
Gambar 4.3 Mensubstitusikan Y	IV-3
Gambar 4.4 <i>Scatterplot</i> dari Data Simulasi Tingkat Kemiskinan	IV-5
Gambar 4.5 <i>Boxplot</i> dari Data Simulasi Tingkat Kemiskinan	IV-5
Gambar 4.6 <i>Scatterplot</i> dari Data Real Tingkat Kemiskinan	IV-19
Gambar 4.7 <i>Boxplot</i> dari Data Real Tingkat Kemiskinan	IV-19

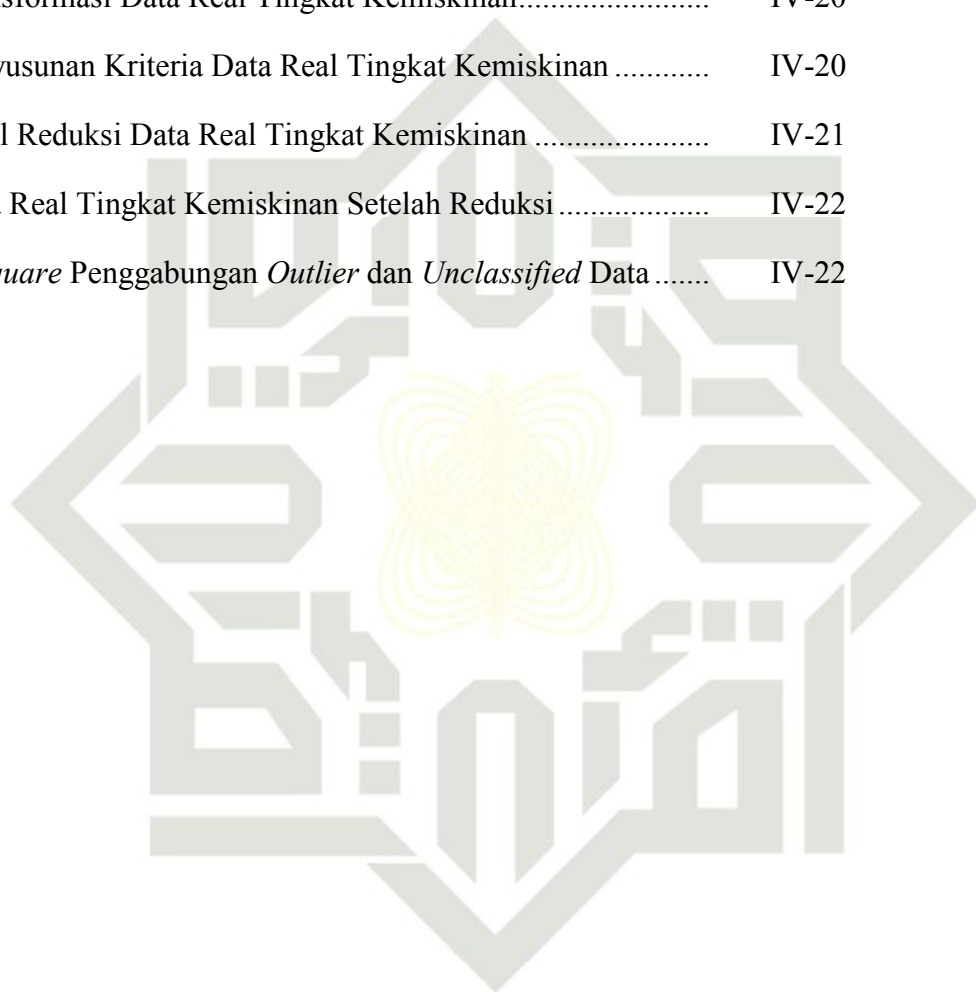
DAFTAR TABEL

Tabel		Halaman
Tabel 2.1	Pengelompokan Data 2 dan 3 Kategori.....	II-5
Tabel 2.2	Pengelompokan Data 4 Kategori dan 5 Kategori.....	II-5
Tabel 2.3	Mengubah Data Numerik ke dalam Bentuk Kategori.....	II-10
Tabel 2.4	Mengelompokkan Data Y	II-10
Tabel 2.5	Hasil Reduksi Data	II-11
Tabel 2.6	Hasil <i>Rough Sets</i>	II-11
Tabel 2.7	Kajian Terkait Penelitian Sebelumnya.....	II-16
Tabel 4.1	Data Simulasi Tingkat Kemiskinan	IV-3
Tabel 4.2	Model Regresi Data Simulasi Sebelum Reduksi	IV-4
Tabel 4.3	<i>Outlier</i> dari 50 Percobaan Sebelum Data Reduksi	IV-6
Tabel 4.4	Transformasi Data Simulasi Tingkat Kemiskinan	IV-7
Tabel 4.5	Penyusunan Kriteria Data Simulasi Tingkat Kemiskinan.....	IV-8
Tabel 4.6	Hasil Reduksi Data Simulasi Tingkat Kemiskinan.....	IV-8
Tabel 4.7	Data Simulasi Tingkat Kemiskinan Setelah Reduksi.....	IV-9
Tabel 4.8	Model Regresi Data Simulasi Tingkat Kemiskinan Setelah Reduksi	IV-10
Tabel 4.9	<i>Unclassified</i> Data dari 50 Percobaan Setelah Data Reduksi..	IV-10
Tabel 4.10	Penggabungan <i>Outlier</i> dan <i>Unclassified</i> Data dari 50 Percobaan.....	IV-12
Tabel 4.11	<i>R-Square</i> dari <i>Outlier</i> , <i>Unclassified</i> Data Serta Penggabungan <i>Outlier</i> dan <i>Unclassified</i> Data.....	IV-14
Tabel 4.12	Uji <i>Paired Samples T Test</i>	IV-14

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 4.13	<i>Paired Samples Statistics</i>	IV-15
Tabel 4.14	<i>Paired Samples Test</i>	IV-15
Tabel 4.15	Data Real Tingkat Kemiskinan.....	IV-17
Tabel 4.16	Deskriptif Data Real Tingkat Kemiskinan.....	IV-18
Tabel 4.17	Transformasi Data Real Tingkat Kemiskinan.....	IV-20
Tabel 4.18	Penyusunan Kriteria Data Real Tingkat Kemiskinan	IV-20
Tabel 4.19	Hasil Reduksi Data Real Tingkat Kemiskinan	IV-21
Tabel 4.20	Data Real Tingkat Kemiskinan Setelah Reduksi	IV-22
Tabel 4.21	<i>R-Square</i> Penggabungan <i>Outlier</i> dan <i>Unclassified Data</i>	IV-22



UIN SUSKA RIAU

BAB I

PENDAHULUAN

Pada Bab I ini akan dibahas mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan, manfaat penelitian dan sistematika penulisan. Selengkapnya akan dijelaskan pada Sub-Bab 1.1-1.6.

1.1 Latar Belakang

Masalah yang sering muncul dalam analisis regresi adalah ditemukannya satu atau beberapa titik data yang berada jauh dari pola data pada umumnya atau yang biasa disebut sebagai *outlier* (pencilan). *Outlier* biasanya timbul karena kesalahan pada penginputan data, kesalahan pengambilan sampel, subjek yang mengerjakan secara asal-asalan ataupun fakta di lapangan memang demikian adanya. *Outlier* ini memang banyak masalahnya jika tidak teliti, salah satunya yaitu *outlier* dapat mempengaruhi hasil estimasi parameter regresi, serta juga dapat menimbulkan pelanggaran terhadap asumsi kenormalan data. *Outlier* dalam analisis regresi dapat menyebabkan sisaan yang besar dari model yang terbentuk dan keragaman data menjadi lebih besar sehingga menyebabkan data tidak homogen (George dan Alan, 2003).

Penelitian yang berkaitan dengan efek *outlier* dalam pengolahan data statistik seperti pada penelitian yang telah dilakukan oleh (Herni, 2002; Ruri, 2008; Abdurakhman, 2002). Selanjutnya penelitian yang dilakukan oleh (Gusti, 2014). Kedua penelitian tersebut membahas tentang pengaruh *outlier* terhadap estimasi parameter regresi. *Outlier* pada model regresi sama halnya dengan *unclassified* data pada model *rough sets*. Dalam model *rough sets* ada istilah yang dikenal dengan *unclassified* data (data eliminasi). *Unclassified* data merupakan suatu tahapan pada model *rough sets* dengan cara menghilangkan data ataupun menghilangkan variabel yang memiliki *missing value*, karena data yang memiliki *missing value* tidak bisa digunakan sebagai alat pengambilan keputusan. Adapun kesamaan antara *outlier* dan *unclassified* data yaitu menghilangkan data yang dianggap tidak memiliki pengaruh yang besar terhadap suatu penelitian.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Model *rough sets* merupakan model terbaru pada data mining untuk mendapatkan aturan-aturan yang singkat dan tepat dari satu tabel (Pawlak, 2002). Model ini juga telah banyak digunakan beberapa peneliti dalam berbagai bidang, seperti penelitian dalam bidang pendidikan telah dilakukan oleh Ilmawati (2014), Syahputra (2015) dan Prajana (2016). Penelitian terkait dalam bidang ekonomi telah dilakukan oleh Herdiyatomoko (2017), Jamaris (2017). Sedangkan penelitian terkait model *rough sets* dalam bidang kesehatan telah dilakukan oleh Teh dkk., (2017) dan Efendi dkk.(2018a). Model *rough sets* memiliki beberapa kelebihan dibandingkan dengan yang lain (Pawlak, 2002), yaitu: menyediakan algoritma yang efisien untuk menemukan pola yang tersembunyi dalam data, menemukan reduksi dari himpunan data, mengevaluasi signifikansi data, mudah untuk dimengerti, menawarkan interpretasi yang mudah dari hasil, dapat digunakan untuk data kualitatif maupun kuantitatif, serta menghasilkan simptom-simptom keputusan dari data.

Disisi lain, ada beberapa atribut atau simptom yang tersedia dalam bentuk data numerik karena ada alat ukurnya, sehingga pendekatan regresi juga sangat memungkinkan untuk menginvestigasi atribut yang signifikan terhadap variabel yang diteliti. Dimotivasi oleh kondisi di atas, maka ada tendensi untuk menggabungkan kedua model yaitu regresi dan *rough sets* dalam penelitian ini. Beberapa penggabungan *rough sets* dan regresi sudah dilakukan oleh peneliti-peneliti sebelumnya (Efendi dan Deris, 2018; Efendi dkk., 2018a, 2018b) yang secara umum diaplikasikan terhadap data-data medik yang sudah memiliki standar klasifikasi (data kualitatif).

Berdasarkan uraian diatas untuk saat ini belum ada penelitian yang membahas tentang hubungan antara *outlier* dan *unclassified data*. Oleh karena itu, penulis tertarik untuk melakukan penelitian dengan menggabungkan *outlier* dan *unclassified data*. Adapun judul yang penulis angkat yaitu **“Efek Penggabungan Outlier dan Unclassified Data pada Pemodelan Rough-Regresi (Studi Kasus: Tiga Variabel Bebas Numerik)”**.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.2 Rumusan Masalah

Dimotivasi oleh penelitian sebelumnya, rumusan masalah dari penelitian ini adalah sebagai berikut: Bagaimana efek dari penggabungan *outlier* dan *unclassified* data dengan menggunakan pemodelan *rough-regresi* ?.

1.3 Batasan Masalah

Dalam membuat penelitian diperlukan batasan-batasan agar tidak menyimpang dari yang telah direncanakan, sehingga tujuan yang sebenarnya dapat dicapai. Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data yang dibangkitkan melalui model regresi yang sudah diverifikasi dengan teknik simulasi data yaitu sebanyak 50 kali percobaan.
2. Model yang digunakan untuk teknik simulasi dalam penelitian ini adalah model regresi linier dengan tiga variabel bebas (*independent variable*) dengan tipe data numerik yang sudah memenuhi asumsi klasik regresi.

1.4 Tujuan Penelitian

Berdasarkan latar belakang, rumusan masalah, dan batasan masalah maka tujuan penelitian ini adalah untuk mengetahui efek dari penggabungan *outlier* dan *unclassified* data dengan menggunakan pemodelan *rough-regresi*.

1.5 Manfaat Penelitian

Ada dua bentuk manfaat dari penelitian ini, yaitu :

- a. Bagi Keilmuan
Pada penelitian ini, diharapkan dapat dijadikan sebagai salah satu sumber informasi bagi pembaca untuk bahan referensi bagi pihak yang membutuhkan. Serta dapat memberikan kontribusi bagi perkembangan Ilmu Pengetahuan dan Teknologi khususnya di bidang Matematika.
- b. Bagi Penulis
Manfaat bagi penulis yaitu: menambah wawasan tentang efek dari penggabungan *outlier* dan *unclassified* data, mengetahui bagaimana cara

1.6

memodelkan data menggunakan *rough*-regresi serta mampu mengaplikasikannya pada data yang lain.

Sistematika Penulisan

Adapun sistematika dalam penulisan proposal Tugas Akhir ini adalah sebagai berikut :

BAB I

PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

BAB II

LANDASAN TEORI

Bab ini menjelaskan hal-hal yang dijadikan sebagai dasar teori untuk mengembangkan penulisan tugas akhir. Konsep dan teori yang relevan yang perlu dijelaskan seperti, konsep regresi, *rough sets*, konsep data *outlier* dan *unclassified* data, konsep data *training* dan *testing* serta konsep metode simulasi.

BAB III

METODOLOGI PENELITIAN

Bab ini menjelaskan tentang bagaimana detail prosedur yang dilakukan penulis untuk mencapai tujuan penelitian, mulai dari membangkitkan data dengan model regresi melalui teknik simulasi sampai memodelkan data tersebut dengan menggunakan metode *rough*-regresi.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang mengenai cara membangkitkan data melalui teknik simulasi sampai memodelkan data tersebut dengan menggunakan metode *rough*-regresi.

BAB V

PENUTUP

Bab ini berisikan kesimpulan yang menjelaskan inti dari seluruh pembahasan beserta saran.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

Untuk mendukung tujuan dan aktivitas penelitian ini, maka pada Bab II ini ada beberapa konsep dan teori yang relevan yang perlu dijelaskan seperti, konsep regresi, *rough sets*, konsep data *outlier* dan *unclassified* data, konsep data *training* dan *testing* serta konsep metode simulasi. Selengkapnya akan dijelaskan pada Sub-Bab 2.1- 2.7.

2.1 Konsep Regresi Linier

Analisis regresi linier merupakan salah satu teknik dalam statistika yang digunakan untuk membentuk model hubungan antara variabel dependen dengan satu atau lebih variabel independen. Dalam analisis regresi dibedakan dua jenis variabel yaitu variabel dependen (variabel terikat) dan variabel independen (variabel bebas). Regresi linier yang terdiri satu variabel dependen dan satu variabel independen disebut regresi linier sederhana (*simple analysis regression*), sedangkan regresi yang terdiri dari satu variabel dependen dan beberapa variabel independen disebut regresi linier berganda (*multiple analysis regression*) (Gujarati, 2006).

Salah satu tujuan dalam analisis regresi adalah untuk mengestimasi nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui. Hasil dari analisis regresi berupa parameter untuk masing-masing variabel independen. Pada umumnya digunakan metode kuadrat terkecil (MKT) atau *Ordinary Least Square* (OLS) untuk mengestimasi koefisien regresi. Metode kuadrat terkecil adalah suatu metode yang digunakan untuk mengestimasi koefisien parameter regresi dengan cara meminimumkan jumlah kuadrat residual. Penggunaan metode ini memerlukan beberapa asumsi klasik yang harus dipenuhi yaitu normalitas, tidak terdapat multikolinearitas, tidak terdapat heteroskedastisitas dan tidak terdapat autokorelasi. Jika asumsi-asumsi klasik dalam metode kuadrat terkecil terpenuhi maka penduga parameter yang diperoleh bersifat *Best Linier Unbiased Estimation* (BLUE) (Gujarati, 2006).

2.1.1 Bentuk Umum Model Regresi Linier Berganda

Model regresi linier berganda adalah hubungan secara linear antara dua atau lebih peubah tak bebas (X_1, X_2, \dots, X_n) dengan peubah bebas (Y). Secara matematis hubungan tersebut dapat dituliskan sebagai berikut (Gujarati, 2006):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + e_i, \quad (2.1)$$

Dari Persamaan (2.1), Y_i merupakan peubah terikat untuk pengamatan ke- i ($i = 1, 2, \dots, n$), X_{1i}, \dots, X_{ki} merupakan peubah bebas untuk pengamatan ke- i , β_0 merupakan *slope* dan β_1, \dots, β_k merupakan *intercept* model, sedangkan e_i merupakan residual untuk pengamatan ke- i diasumsikan berdistribusi normal yang saling bebas dengan rata-rata 0 (nol) dan variansi σ^2 .

Persamaan regresi linier berganda dengan k variabel bebas memiliki parameter β_0 sebagai konstanta dan koefisien regresi $\beta_1, \beta_2, \beta_3, \dots, \beta_k$. Menentukan nilai $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_k$ dapat dicari secara manual menggunakan prinsip metode kuadrat terkecil atau *ordinary least square* (OLS) yaitu dengan meminimumkan nilai (Sembiring, 2003):

$$J = \sum \varepsilon_i^2 = \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki})^2 \quad (2.2)$$

Langkah-langkah untuk mendapatkan nilai $\beta_0, \beta_1, \beta_2, \beta_3, \dots$, dan β_k adalah:

1. Menentukan fungsi-fungsi turunan pertama parsial dari jumlah J terhadap masing-masing nilai $\beta_0, \beta_1, \beta_2, \beta_3, \dots$, dan β_k lalu disamakan dengan nol, sehingga diperoleh Turunan pertama dari J terhadap β_0 menjadi:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki}) = 0$$

Turunan pertama dari J terhadap β_1 menjadi:

$$\frac{\partial J}{\partial \beta_1} = -2 \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki})(X_{1i}) = 0$$

Turunan pertama dari J terhadap β_2 menjadi:

$$\frac{\partial J}{\partial \beta_2} = -2 \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki})(X_{2i}) = 0$$

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

Turunan pertama dari J terhadap β_3 menjadi:

$$\frac{\partial J}{\partial \beta_3} = -2 \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki})(X_{3i}) = 0$$

Turunan pertama dari J terhadap β_k menjadi:

$$\frac{\partial J}{\partial \beta_k} = -2 \sum (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_k X_{ki})(X_{ki}) = 0$$

- Menyederhanakan hasil turunan parsial tersebut lalu mengganti koefisien regresi dengan penaksirannya, diperoleh:

$$\begin{aligned} \sum Y_i &= n\beta_0 + \beta_1 \sum X_{1i} + \beta_2 \sum X_{2i} + \dots + \beta_k \sum X_{ki} \\ \sum Y_i X_{1i} &= \beta_0 \sum X_{1i} + \beta_1 \sum X_{1i}^2 + \beta_2 \sum X_{1i} X_{2i} + \dots + \beta_k \sum X_{1i} X_{ki} \\ \sum Y_i X_{2i} &= \beta_0 \sum X_{2i} + \beta_1 \sum X_{1i} X_{2i} + \beta_2 \sum X_{2i}^2 + \dots + \beta_k \sum X_{2i} X_{ki} \\ \sum Y_i X_{3i} &= \beta_0 \sum X_{3i} + \beta_1 \sum X_{1i} X_{3i} + \beta_2 \sum X_{2i} X_{3i} + \dots + \beta_k \sum X_{3i} X_{ki} \\ \sum Y_i X_{ki} &= \beta_0 \sum X_{ki} + \beta_1 \sum X_{1i} X_{ki} + \beta_2 \sum X_{2i} X_{ki} + \dots + \beta_k \sum X_{ki}^2 \end{aligned}$$

- Untuk mempermudah perhitungan, persamaan tersebut dapat dirubah kedalam bentuk matriks, sehingga bentuknya menjadi:

$$\begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \dots & \sum X_{1i} X_{ki} \\ \sum X_{2i} & \sum X_{1i} X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ki} & \sum X_{1i} X_{ki} & \sum X_{2i} X_{ki} & \dots & \sum X_{ki}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1i} \\ \sum Y_i X_{2i} \\ \vdots \\ \sum Y_i X_{ki} \end{bmatrix} \quad (2.3)$$

Dengan:

$$A = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \dots & \sum X_{1i} X_{ki} \\ \sum X_{2i} & \sum X_{1i} X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ki} & \sum X_{1i} X_{ki} & \sum X_{2i} X_{ki} & \dots & \sum X_{ki}^2 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \text{dan} \quad J = \begin{bmatrix} \sum Y_i \\ \sum Y_i X_{1i} \\ \sum Y_i X_{2i} \\ \vdots \\ \sum Y_i X_{ki} \end{bmatrix}$$

Sehingga persamaan matrik (2.3) dapat dibentuk menjadi:

$$AB = J \quad (2.4)$$

- Menentukan matrik B yaitu dengan menentukan A^{-1} terlebih dahulu untuk mempermudah, maka pada persamaan (2.4) kedua ruas sama-sama dikalikan dengan A^{-1} sehingga:

$$\begin{aligned} A^{-1}AB &= A^{-1}J \\ B &= A^{-1}J \end{aligned} \quad (2.5)$$

5. Setelah matrik B diperoleh maka nilai $\beta_0, \beta_1, \beta_2, \beta_3, \dots$, dan β_k sudah diketahui sehingga dapat dibuat sebuah model regresi linier berganda.

2.1 Asumsi-asumsi Model Regresi Linier Berganda

Menurut Gujarati (2003) asumsi-asumsi pada model regresi linier berganda adalah sebagai berikut:

- Model regresinya adalah linier dalam parameter.
- Nilai rata-rata dari *error* adalah nol.
- Variansi dari *error* adalah konstan (homoskedastik).
- Tidak terjadi autokorelasi pada *error*.
- Tidak terjadi multikolinieritas pada variabel bebas.
- Error* berdistribusi normal.

Berikut adalah asumsi-asumsi pada regresi linier berganda yang harus terpenuhi sebelum membuat model:

A. Uji Normalitas

Uji normalitas adalah sebuah uji yang dilakukan dengan tujuan untuk menilai sebaran data pada sebuah kelompok data atau variabel berdistribusi normal atau tidak. Uji normalitas berguna untuk menentukan data yang telah dikumpulkan berdistribusi normal atau diambil dari populasi normal. (Utama, 2009).

Uji normalitas dapat dilakukan dengan beberapa metode, dari yang paling sederhana sampai yang canggih. Setiap metode memiliki keunggulan dan kelemahan masing-masing. Beberapa metode yang dapat dilakukan untuk melakukan pemeriksaan kenormalan diantaranya yaitu dengan melihat penyebaran data pada sumbu diagonal pada grafik *normal P-P plot of regression standardized residual* atau dengan uji *one sample kolmogorov-smirnov*.

Uji normalitas data ini berfungsi ketika kita mengkriteriakan/mengkategorikan data pada distribusi frekuensi (Santoso, 2001). Selengkapnya disajikan pada Tabel 2.1.

Tabel 2.1 Pengelompokan Data 2 dan 3 Kategori (Santoso, 2001)

Klasifikasi		Data Berdistribusi Normal	Data Tidak Berdistribusi Normal
2 Kriteria	Tinggi	$\text{Data} \geq \text{Mean}$	$\text{Data} \geq \text{Median}$
	Rendah	$\text{Data} < \text{Mean}$	$\text{Data} < \text{Median}$
3 Kriteria	Tinggi	$\text{Data} > \text{Mean} + \text{SD}$	$\text{Data} > Q_3$
	Sedang	$\text{Mean} - \text{SD} \leq \text{Data} \leq \text{Mean} + \text{SD}$	$Q_1 \leq \text{Data} \leq Q_3$
	Rendah	$\text{Data} < \text{Mean} - \text{SD}$	$\text{Data} < Q_1$

Berdasarkan Tabel 2.1 klasifikasi data bisa dilakukan berdasarkan kenormalan. Kriteria terbagi atas 2 dan 3. Jika data berdistribusi normal untuk 2 kriteria menggunakan pendekatan *mean*, sedangkan untuk data yang tidak berdistribusi normal menggunakan pendekatan median. Selanjutnya klasifikasi untuk 3 kriteria yang memiliki distribusi normal menggunakan pendekatan *mean* dan standar deviasi (SD), sedangkan untuk data yang tidak berdistribusi normal menggunakan pendekatan kuartil (Q_1, Q_2, Q_3). Untuk mengkategorikan hasil pengukuran menjadi 4 dan 5 kategori, pedoman yang bisa digunakan adalah

Tabel 2.2 Pengelompokan Data 4 Kategori (Cooper dan Schindler, 2006) dan 5 Kategori (Azwar, 2012)

Klasifikasi		Kelas Interval
4 Kriteria	Rendah	$\text{Data} < Q_1$
	Sedang	$Q_1 < \text{Data} < Q_2$
	Tinggi	$Q_2 < \text{Data} < Q_3$
	Sangat Tinggi	$\text{Data} > Q_3$
5 Kriteria	Sangat Rendah	$\text{Data} \leq M - 1,5\text{SD}$
	Rendah	$M - 1,5\text{SD} < \text{Data} \leq M - 0,5\text{SD}$
	Sedang	$M - 0,5\text{SD} < \text{Data} \leq M + 0,5\text{SD}$
	Tinggi	$M + 0,5\text{SD} < \text{Data} \leq M + 1,5\text{SD}$
	Sangat Tinggi	$M + 1,5\text{SD} < \text{Data}$

Berdasarkan Tabel 2.2 klasifikasi data terbagi atas 4 dan 5 kriteria. Klasifikasi untuk 4 kriteria menggunakan pendekatan kuartil (Q_1, Q_2, Q_3). Dan klasifikasi untuk 5 kriteria menggunakan pendekatan *mean* dan standar deviasi (SD).

B. Uji Multikolinearitas

Uji Multikolinearitas adalah terjadinya hubungan linier antara variabel bebas dalam suatu model regresi linier berganda (Gujarati, 2003). Uji multikolinearitas bertujuan untuk mendeteksi ada atau tidaknya multikolinearitas di dalam model regresi dapat diketahui dengan melihat *Variance Inflation Factor* (*VIF*) dan *tolerance* yang mana jika nilai *VIF* lebih kecil dari 10 dan *tolerance* lebih besar dari 0,10, maka model regresi bebas multikolinearitas.

C. Uji Heteroskedastisitas

Uji heteroskedastisitas adalah variansi dari error model regresi tidak konstan atau variansi antar error yang satu dengan error yang lain berbeda (Widarjono, 2007). Gejala ini dapat mengurangi akurasi hasil analisis regresi. Ada tidaknya heteroskedastisitas dapat diketahui dengan melihat grafik *Scatterplot* pada output SPSS. Dasar pengambilan keputusannya adalah jika titik-titik pada grafik membentuk pola tertentu (bergelombang, melebar atau menyempit) maka dapat diindikasikan telah terjadi heteroskedastisitas, sedangkan jika titik-titik pada grafik tidak membentuk pola serta titik-titik menyebar di atas dan dibawah angka 0 pada sumbu *Y* maka dapat diindikasikan tidak terjadi heteroskedastisitas.

D. Uji Autokorelasi

Uji Autokorelasi adalah terjadinya korelasi antara satu variabel error dengan variabel error yang lain. Autokorelasi seringkali terjadi pada data *time series* dan dapat juga terjadi pada data *cross section* (Widarjono, 2007). Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pengganggu pada periode $t-1$ (sebelumnya). Autokorelasi muncul karena observasi yang berurutan sepanjang waktu berkaitan satu sama lainnya. Model regresi yang baik adalah regresi yang bebas dari autokorelasi. Dasar pengambilan keputusannya adalah dengan cara membandingkan DW (Durbin-Watson) hasil dengan DW tabel.

- a) Jika $d < d_l$ maka tolak H_0
- b) Jika $d_l < d < d_u$ maka tidak dapat kesimpulan

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- c) Jika $du < d < 4 - du$ maka tidak ada autokorelasi
- d) Jika $4 - du < d < 4 - dl$ maka tidak ada autokorelasi
- e) Jika $d > 4 - dl$ maka tolak H_0

2.1.3 Pengujian Parameter Regresi Linier Berganda

Pengujian parameter ini bertujuan untuk mengetahui ada atau tidaknya pengaruh variabel bebas terhadap variabel tidak bebas, baik secara serentak maupun secara parsial.

A. Pengujian Parameter Secara Serentak (Uji- F)

Uji- F ini dilakukan untuk mengetahui apakah variabel-variabel bebas yang digunakan berpengaruh secara simultan terhadap variabel terikat. Proses pengujian dapat dilakukan dengan pendekatan analisis variansi (Sugiyono, 2009).

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1 : \text{Minimal ada satu } \beta_k \neq 0, k = 0, 1, 2, 3, \dots, p$$

Rumus statistik uji dengan analisis variansi adalah sebagai berikut:

$$F_{hit} = \frac{RKR}{RKS} \quad (2.6)$$

Dari Persamaan (2.2), RKR adalah rata-rata kuadrat regresi, dan RKS adalah rata-rata kuadrat sisa (dapat diperoleh dari Tabel Analisis Variansi). Teknik Pengambilan keputusan pada uji ini adalah sebagai berikut:

H_0 diterima jika $F_{hitung} < F_{tabel}$, artinya tidak terdapat pengaruh semua variabel bebas terhadap variabel terikat secara bersama-sama.

H_0 ditolak jika $F_{hitung} > F_{tabel}$, artinya terdapat pengaruh semua variabel bebas terhadap variabel terikat secara bersama-sama.

B. Pengujian Parameter Secara Parsial (Uji- t)

Selain Uji- F perlu juga dilakukan Uji- t yang digunakan untuk mengetahui apakah masing-masing variabel bebas secara sendiri-sendiri mempunyai pengaruh secara signifikan terhadap variabel terikat. Dengan kata lain untuk mengetahui apakah masing-masing variabel bebas dapat menjelaskan

perubahan yang terjadi pada variabel terikat secara nyata. Uji t digunakan untuk membuat keputusan apakah hipotesis terbukti atau tidak, di mana tingkat signifikan yang digunakan yaitu 5%. Adapun nilai peubah acak T yang mempunyai sebaran t dengan $v = n - 1$ derajat bebas adalah sebagai berikut (Sudiyono, 2009):

$$t = \frac{\bar{x} - \mu}{s^2 / \sqrt{n}} \quad (2.7)$$

Dimana \bar{x} dan s^2 masing-masing adalah nilai tengah dan ragam suatu contoh acak berukuran n yang diambil dari suatu populasi normal dengan nilai tengah μ dan ragam σ^2 . Apabila harga koefisien t yang digunakan sebagai ukuran, maka harga koefisien tersebut harus dibandingkan dengan dengan harga t tabel untuk tingkat signifikan yang telah ditetapkan dengan df yang sesuai.

Pengambilan keputusan:

H_0 diterima jika $-t_{tabel} < t_{hitung} < t_{tabel}$ artinya suatu variabel bebas tidak mempengaruhi variabel terikat dengan variabel bebas lainnya sebagai pengontrol.

H_0 ditolak jika t_{hitung} berada diluar selang $(-t_{tabel}, t_{tabel})$, artinya suatu variabel bebas mempengaruhi variabel terikat dengan variabel bebas lainnya sebagai pengontrol.

2.2 Konsep Rough Sets

Konsep ini diperkenalkan pertama kali oleh Zdzislaw Pawlak pada tahun 1982 untuk menganalisis data-data yang penuh ketidakpastian serta dalam bentuk kategori atau kriteria dengan menggunakan pendekatan teori himpunan (Pawlak, 1982). Selanjutnya hasil analisis tersebut digunakan untuk pembentukan *rule* dalam sistem pengambil keputusan (SPK). Ada beberapa komponen penting yang ada pada konsep ini, seperti; *information systems*, *indiscernibility relation*, *set approximations*, *rough clustering*, dan sebagainya. Sebuah information system $S = (U, \Omega, V_q, f_q)$ terdiri dari (Polkowski dkk., 1998; Rissino dan Tores, 2009):

U merupakan sebuah himpunan universal dan tidak kosong;

Ω merupakan himpunan atribut dan tidak kosong;

$\Omega = C \cup D$, dimana C merupakan himpunan atribut-atribut bersyarat (*conditional attributes*) dan D merupakan himpunan atribut keputusan (*decision attribute*) dan kedua himpunan merupakan himpunan berhingga; untuk setiap $q \in \Omega$, V_q dikenal sebagai domain q ; f_q : sebuah information system $f_q: U \rightarrow V_q$.

Objek atau elemen pada himpunan atribut bisa berupa, kasus, proses, pasien, mahasiswa, perusahaan, dsb. Sedangkan atribut bisa berupa, simptom, faktor, variabel, karakteristik dari sebuah informasi, fitur, dsb.

Tujuan analisis *rough sets* adalah untuk mendapatkan *rule* yang klasifikasi setelah dilakukan pengumpulan data. *Rule* disini sudah diklasifikasikan setelah mendapatkan *reduct*. Sebagai contoh, pasien yang menderita penyakit flu, memiliki gejala yang sama tetapi tak terlihat dan dapat dianggap sebagai unit penyakit pengetahuan medis. Pengetahuan medis ini disebut set dasar (konsep).

Rough sets menentukan teorinya menggunakan perkiraan, yaitu yang ditentukan oleh fungsi keanggotaan. *Rough sets* bisa juga menentukan teorinya tanpa menggunakan perkiraan. Karena fungsi keanggotaan bukanlah konsep primitif dalam pendekatan yang dalam hal ini kedua defenisi tidak setara. Fungsi keanggotaan merupakan pemetaan titik-titik yang didapat dari himpunan *fuzzy* kedalam keanggotaan yang memiliki interval antara 0 sampai dengan 1. Salah satu cara untuk mendapatkan nilai keanggotaan adalah dengan pendekatan fungsi.

Menurut Hartama dan Hartono (2016) bahwa, tahapan di dalam penggunaan algoritma *rough sets* ini sebagai berikut:

- Data selection* (pemilihan data yang akan digunakan).
- Pembentukan *decision system* yang berisikan atribut kondisi dan atribut keputusan.
- Pembentukan *equivalence class*, yaitu dengan menghilangkan data yang berulang.
- Pembentukan *discernibility matrix modulo D*, yaitu matriks yang berisikan perbandingan antar data yang berbeda atribut kondisi dan atribut keputusan.
- Menghasilkan *reduct* dengan menggunakan aljabar boolean.
- Menghasilkan *rule* (pengetahuan).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Untuk menunjukkan kerja dari algoritma *rough sets* yang dikemukakan oleh Hartama dan Hartono diatas, maka penulis mengambil sebuah contoh yaitu pada penelitian metode prediksi fertilitas (Efendi dkk., 2018)

- a. Mengubah data numerik menjadi data kategori

Data real yang diperoleh dari hasil penelitian diubah ke dalam bentuk data kategori. Secara detail ditampilkan pada Tabel 2.3 sebagai berikut:

Tabel 2.3 Mengubah Data Numerik ke dalam Bentuk Kategori

Responden	Y	X_1	X_2	X_3	X_4	X_5
A1	Banyak	Banyak	Tinggi	Normal	Sedang	Iya
A2	Sedikit	Sederhana	Tinggi	Tua	Cepat	Iya
A3	Sederhana	Sederhana	Rendah	Normal	Lama	Iya
A65	Sederhana	Banyak	Tinggi	Tua	Cepat	Iya

- b. Mengelompokkan Data Berdasarkan Variabel Y

Setelah mengubah data ke bentuk kategori, selanjutnya yaitu mengelompokkan data berdasarkan fertilitas Y dengan mengurutkan variabel Y dari yang tertinggi ke rendah. Secara detail ditampilkan pada Tabel 2.4 sebagai berikut:

Tabel 2.4 Mengelompokkan Data Y

Responden	Y	X_1	X_2	X_3	X_4	X_5
A1	Banyak	Banyak	Tinggi	Normal	Sedang	Iya
A2	Banyak	Banyak	Tinggi	Normal	Sedang	Iya
A3	Banyak	Banyak	Sedang	Muda	Lama	Tidak
A65	Sedikit	Sederhana	Sedang	Normal	Cepat	Tidak

Dari tabel 2.4 di atas dapat dilihat bahwa variabel terikat (Y) sudah terurut dari yang tertinggi ke yang terendah.

- c. Mengeliminasi atau mereduksi data

Langkah selanjutnya yaitu mengeliminasi data yang sama dengan berdasarkan pada variabel Y . Secara detail ditampilkan pada Tabel 2.5 sebagai berikut:

Tabel 2.5 Hasil Reduksi Data

Responden	Y	X_1	X_2	X_3	X_4	X_5
A1	Banyak	Banyak	Tinggi	Normal	Sedang	Iya
A2	Banyak	Banyak	Tinggi	Normal	Sedang	Iya
A61	Sederhana	Sederhana	Rendah	Muda	Sedang	Tidak

Dari tabel 2.5 dapat disimpulkan bahwa dari 65 responden terdapat 9 responden yang dihilangkan karena data yang dimiliki terjadi missing value sehingga data tidak bisa digunakan sebagai alat pengambilan keputusan.

d. Membuat *rule*

Pada bagian ini akan ditampilkan *rule* yang dibuat berdasarkan data yang dieliminasi, dengan mengambil *rule* yang paling dominan untuk dijadikan acuan yakni sebagai berikut:

Rule 1: Jika X_1 “Banyak”, X_2 “Tinggi”, X_3 “Normal”, X_4 “Sedang”, X_5 “Iya”, maka Y “Banyak”.

Hasil *rule* di atas adalah informasi baru yang didapatkan sehingga berdasarkan informasi di atas dapat dijadikan acuan dan mempermudah dalam menggunakan metode *rough sets*.

e. Membuat *rough sets* dengan menggunakan *rule*

Setelah mendapatkan *rule* langkah selanjutnya yaitu membuat *rough sets* dengan menggunakan *rule*, dimana data kategori awal dicari *rough sets* nya menggunakan *rule* yang sudah diperoleh dari langkah sebelumnya. Berikut ini hasil *rough sets* yang diperoleh:

Tabel 2.6 Hasil Rough Sets

Respo n	Y	X_1	X_2	X_3	X_4	X_5	<i>Rough sets</i>
A1	Banyak	Banyak	Tinggi	Normal	Sedang	Iya	Banyak
A2	Banyak	Banyak	Tinggi	Normal	Sedang	Iya	Banyak
A65	Sedikit	Sederhana	Sedang	Normal	Cepat	Tidak	Sederhana

Dari tabel 2.6 di atas *rough sets* dari 65 responden, yang mana data kategori yang diperoleh hanya ada 2 kategori yaitu “Banyak” pada responden A1, A2 dan kategori sederhana pada responden A3-A65.

2.3 Konsep Data *Outlier*

Outlier adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim baik untuk sebuah variabel tunggal atau kombinasi (Ghozali, 2011:41). Menurut (Ghozali, 2011:41) ada empat penyebab timbulnya data *outlier* (1) kesalahan dalam mengentri data, (2) gagal menspesifikasi adanya *missing value* dalam program komputer, (3) *outlier* bukan merupakan anggota populasi yang kita ambil sebagai sampel, tetapi (4) *outlier* berasal dari populasi yang kita ambil sebagai sampel, tetapi distribusi dari variabel dalam populasi tersebut memiliki nilai ekstrim dan tidak berdistribusi secara normal. Oleh karena itu apabila dalam suatu data pengamatan terdapat *outlier*, maka untuk mencari solusinya tidak diperkenankan menggunakan metode *Least Square Estimate* (LS) karena metode LS didasarkan pada asumsi bahwa error dari model yang dihasilkan harus berdistribusi normal.

Error yang merupakan *outlier* adalah yang nilai mutlaknya jauh lebih besar daripada error lainnya dan bisa jadi terletak tiga atau empat kali simpangan baku atau lebih jauh lagi dari rata-rata errornya. *Outlier* merupakan suatu keganjilan dan menandakan suatu titik data yang sama sekali tidak tipikal dibandingkan data lainnya (Draper dan Smith, 1998). Sebagaimana yang telah dikemukakan Soeharti (2007) bahwa Ferguson mendefinisikan *outlier* sebagai suatu pengamatan yang menyimpang dari sekumpulan pengamatan yang lain. Barnett mendefinisikan *outlier* adalah pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat.

Adakalanya *outlier* memberikan informasi yang tidak bisa diberikan oleh titik lainnya, misalnya karena *outlier* timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. *Outlier* merupakan nilai ekstrim dari suatu pengamatan. Seperti yang diketahui bahwa pada analisis regresi terdapat suatu variabel dependen dan independen yang digambarkan dalam *scatterplot* sebagai arah x dan y , oleh karena itu kemungkinan keberadaan *outlier* bisa terjadi pada arah x atau y dan bahkan bisa terjadi pada arah x dan y sekaligus.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Apabila data *outlier* terdapat pada arah x , maka akan memberikan pengaruh yang sangat besar pada estimator metode LS karena *outlier* pada arah x akan membalikkan garis ideal (LS). Oleh karena itu, *outlier* pada arah x disebut dengan titik *leverage*. Sedangkan apabila data *outlier* itu terletak pada arah y , maka akan memberikan nilai residual r yang sangat besar (negatif atau positif) sehingga tidak menunjukkan garis LS.

Untuk mengetahui ada atau tidaknya data *outlier* dapat dicari dengan menggunakan diagram pencar dari variabel independen maupun dependen atau dapat juga dicari dengan pemeriksaan visual dari data mentahnya, akan tetapi apabila dalam suatu kasus ketika terdapat lebih dari dua variabel independen, maka akan sulit ditemukan data yang mengandung *outlier*. Oleh karena itu, cara mendeteksi terhadap *outlier* dapat dilakukan dengan menentukan nilai batas yang akan dikategorikan sebagai data *outlier* yaitu dengan cara mengkonversi nilai data ke dalam skor *standardized* atau yang biasa disebut *z-score* (Ghozali, 2011:41). Untuk kasus sampel kecil (kurang dari 80) maka standar skor dengan nilai $\geq 2,5$ dinyatakan *outlier*.

A. Dampak *Outlier* (pencilan)

Keberadaan data pencilan akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut (Paludi, 2009):

- Residual yang besar dari model yang terbentuk atau $E[e] \neq 0$.
- Varians pada data tersebut menjadi lebih besar.
- Taksiran interval memiliki rentang yang lebar.

B. Identifikasi *Outlier* (pencilan)

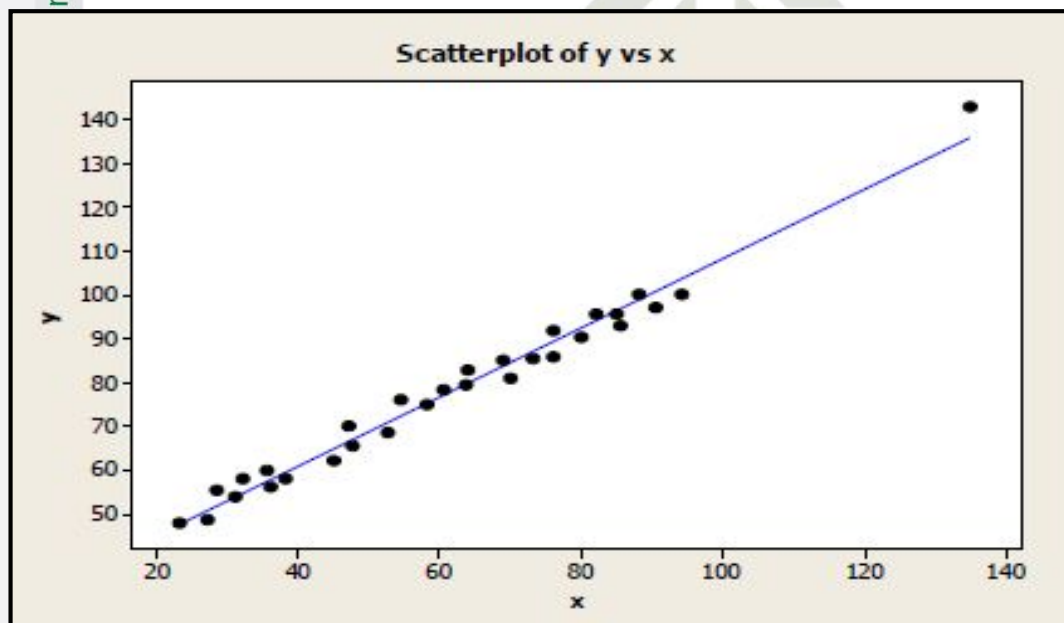
Suatu data mungkin letaknya terpencil, tapi bila pengaruhnya terhadap koefisien kecil maka kita tidak perlu memberi perhatian besar padanya. Makin besar ukuran sampel n makin kecil pengaruh suatu titik data. Pengaruh suatu data mungkin merupakan pencilan bila menggunakan suatu model, tapi tidak bila model lain yang digunakan.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dalam statistik ruang, data pencilan harus dilihat terhadap posisi dan sebaran data yang lainnya sehingga akan dievaluasi apakah data pencilan tersebut perlu dihilangkan atau tidak. Terdapat beberapa metode untuk menentukan batasan pencilan dalam sebuah analisis, yaitu:

1. Metode Grafis

Untuk melihat apakah terdapat pencilan pada data, dapat dilakukan dengan memplot antara data dengan observasi ke- i ($i=1,2,3, \dots, n$) seperti gambar berikut:



Gambar 2.1 Scatter-Plot dari Data dengan Observasi ke- i

Dari contoh di atas terdapat salah satu data, yakni observasi yang letaknya jauh dari pola data keseluruhan diidentifikasi merupakan pencilan. Selain melalui *scatter-plot* di atas, jika sudah didapatkan model regresi maka dapat dilakukan dengan cara memplot antara residual (e) dengan nilai prediksi $y(\hat{y})$. Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan maka hal ini mengindikasikan adanya pencilan. Kelemahan dari metode ini adalah keputusan bahwa suatu data merupakan pencilan sangat bergantung pada *judgement* peneliti, karena hanya mengandalkan visualisasi grafis, untuk itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut. Dalam rangka meminimumkan kesalahan

2. *Boxplot*

Hak Cipta Dilindungi Undang-Undang

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

-
- ☆ Nilai Ekstrim
- Pencilan
- 1,5 R
- 3R
- Q3
- Q2
- Q1
- $R = Q3 - Q1$
- 1,5 R
- 3R
- Pencilan
- ☆ Nilai Ekstrim
- Batas bukan pencilan

Gambar 2.2 Skema Identifikasi Pencilan Menggunakan IQR atau *Boxplot*

C. Tindakan Terhadap *Outlier* (pencilan)

Tujuan analisis data pencilan bukan hanya mencari pola umum data tapi juga mencari data yang mana saja yang tidak mengikuti pola umum. Bila ternyata hasil identifikasi menunjukkan adanya pencilan, maka yang dapat dilakukan adalah membuang atau menghilangkan data pengamatan tersebut, jika tidak memberikan pengaruh setelah dilakukan pengujian. Karena bagaimanapun juga keberadaan data pencilan mengganggu proses analisis. Untuk menanggulangi pencilan data, yaitu dengan mengeluarkan atau membuang observasi ke- i pada data yang diduga merupakan pencilan. Kemudian dilakukan pengujian kembali untuk mendeteksi terdapat atau tidaknya pencilan pada data sampai tidak terdapat lagi pencilan pada data tersebut.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4 Konsep *Unclassified* Data

Unclassified data (data eliminasi) adalah suatu tahapan dalam metode *rough sets* dengan cara menghilangkan data ataupun menghilangkan variabel yang tidak konsisten dalam data set atau data yang tidak memiliki klasifikasi tertentu. Sehingga diperoleh data yang benar-benar bagus untuk dijadikan sebagai alat dalam pengambilan keputusan.

2.5 Konsep Data *Training* dan *Testing*

Data dibagi menjadi dua bagian, yaitu data *training* dan *testing*. Data *training* digunakan untuk membentuk model dalam analisis. Sedangkan data *testing* digunakan untuk menguji ketepatan prediksi dari model yang terbentuk. Jumlah data *training* dan *testing* masing-masing dipilih 80% dan 20% dari keseluruhan data (Zufa dkk, 2017).

2.6 Konsep Metode Simulasi

Simulasi ialah suatu metode untuk melaksanakan percobaan dengan menggunakan model dari suatu sistem nyata (Siagian, 1987). Simulasi merupakan suatu model pengambilan keputusan dengan mencontoh atau mempergunakan gambaran sebenarnya dari suatu sistem kehidupan nyata tanpa harus mengalaminya pada keadaan yang sesungguhnya (Hasan, 2002).

2.7 Kajian Penelitian Sebelumnya

Dalam penelitian ini penulis mengambil beberapa jurnal rujukan mengenai *outlier*, *unclassified* data dan pemodelan *rough*-regresi. Selengkapnya disajikan melalui Tabel 2.7.

Tabel 2.7 Kajian Terkait Penelitian Sebelumnya

No	Nama Peneliti dan Tahun	Model dan Data yang digunakan	Deskriptif dan Kajian
1	Paludi, 2009	Metode grafis dan boxplot	Pada kajian ini menjelaskan tentang bagaimana cara mengidentifikasi <i>outlier</i> dan melihat pengaruh keberadaan <i>outlier</i> .

Hak Cipta Dilindungi Undang-Undang

2	Efendi dkk., 2018	Regresi, <i>Rough sets</i> , <i>Rough-regresi</i> data fertilitas	Pada kajian ini menjelaskan tentang variabel yang mempengaruhi fertilitas.
3	Efendi dkk., 2017	<i>Rough-regresi</i> dan literasi keuangan mahasiswa	Pada kajian ini menjelaskan penentuan kriteria pengetahuan keuangan pada mahasiswa menggunakan <i>rough-regresi</i> .
4	Efendi dan Deris., 2018	<i>Rough-regresi</i> dan pasien kolesterol	Pada kajian ini dijelaskan model pengambilan keputusan untuk pasien kolesterol dan kriteria dominan dengan menggunakan <i>rough-regresi</i>
5	Efendi dkk., 2018a	<i>Rough-regresi</i> dan diagnosa data medis	Pada kajian ini dibahas diagnosa medis dan prediksi menggunakan model <i>rough-regresi</i> .
6	Efendi dkk., 2018b	<i>Rough-regresi</i> dan pasien flu, diabetes, cikungnya, kanker	Pada kajian ini dibahas pentingnya peran data eliminasi untuk meningkatkan akurasi peramalan penyakit pasien.

Berdasarkan Tabel 2.7, Kajian 1 dijadikan sebagai sumber rujukan dalam mengidentifikasi keberadaan *outlier* namun pada kajian 1 ini belum membahas secara detail mengenai pengaruh *outlier* terhadap *R-Square*. Oleh karena itu, dalam penelitian tugas akhir ini akan dibahas lebih rinci mengenai efek yang ditimbulkan dari *outlier* terkhusus dengan mengamati *R-Square* dari setiap percobaan. Pada Kajian 2 membahas tentang *unclassified* data (data eliminasi) namun masalahnya kajian 2 ini tidak mempertimbangkan data *outlier* dulu. Sehingga jika antara *outlier* dan *unclassified* data sama-sama ditinjau terlebih dahulu, maka dapat dibuat kesimpulan langkah yang lebih bagus. Melalui penelitian ini penulis ingin menggabungkan apakah dengan adanya penggabungan *outlier* dan *unclassified* data dapat memperoleh *R-Square* yang baik. Begitu juga Kajian 3 sampai kajian 6 yang dijadikan rujukan dalam membahas metode yang digunakan untuk mengetahui efek dari penggabungan *outlier* dan *unclassified* data.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III METODOLOGI PENELITIAN

Pada Bab III ini akan dijelaskan bagaimana detail prosedur yang dilakukan penulis untuk mencapai tujuan penelitian pada bagian (1.4) dengan menggunakan model *rough-regresi*. Selengkapnya akan dibahas pada Sub-Bab 3.1-3.6.

3.1 Tahap Persiapan Data

Ada beberapa tahapan yang perlu dipersiapkan sebelum melakukan uji efek penggabungan *outlier* dan *unclassified* data, seperti:

- Jenis data : Data yang dibangkitkan melalui *software* Minitab.
- Jumlah data : Data yang digunakan sebanyak 63 data.
- Pengumpulan Data : Metode pengumpulan data dilakukan dengan teknik simulasi.
- Jumlah percobaan : Data yang disimulasikan sebanyak 50 percobaan.
- Jumlah variabel : Penelitian ini menggunakan tiga variabel bebas (*independent variable*) dan satu variabel terikat (*dependent variable*).

3.2 Tahap Prosedur Simulasi

Adapun langkah-langkah yang dilakukan pada tahap prosedur simulasi regresi ini adalah sebagai berikut:

- Menentukan model regresi. Model regresi dirujuk dari Jurnal Ekonomi dan Bisnis yang telah diverifikasi (model sudah memenuhi asumsi-asumsi klasik pada regresi linier).

$$\hat{Y} = 15,154 + 0,259 X_1 + 0,283 X_2 + 0,167 X_3 \quad (3.1)$$

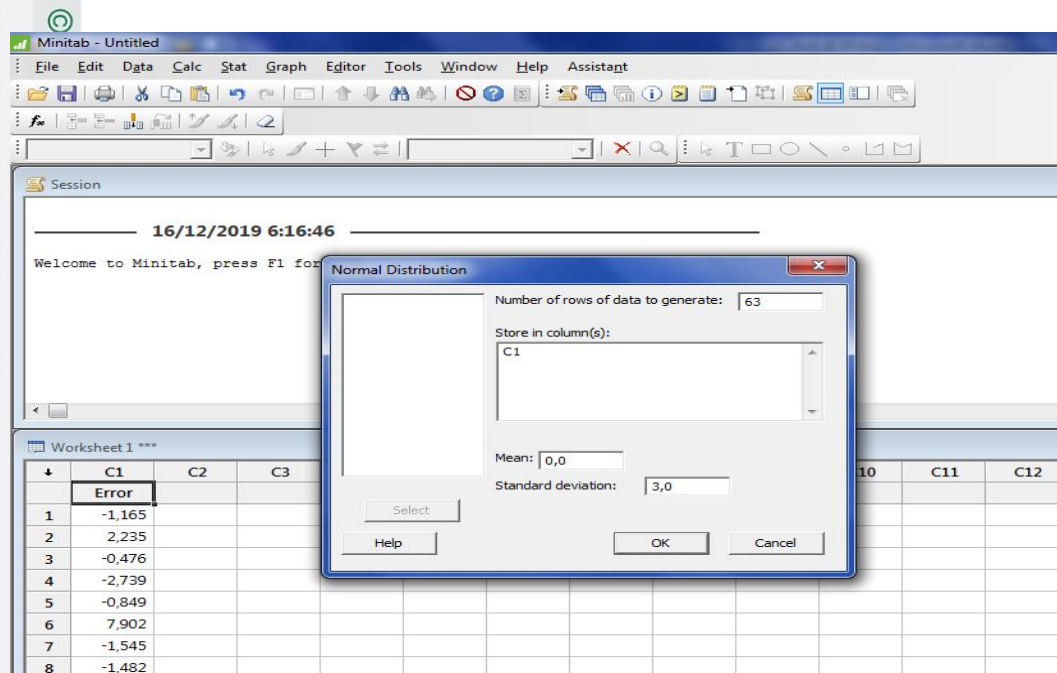
- Membangkitkan nilai $error \sim N(0,1)$. *Error* yang dibangkitkan terdistribusi secara normal. Adapun langkah-langkah membangkitkan *error* dengan menggunakan *software* Minitab yaitu dengan cara klik Calc > Random data < Normal > pada kotak dialog “Number of rows of data to generate” tulis jumlah yang akan dimasukkan (dalam penelitian ini *error* dibangkitkan sebanyak 63) dan pada kolom “store in column (s)” masukkan *error*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



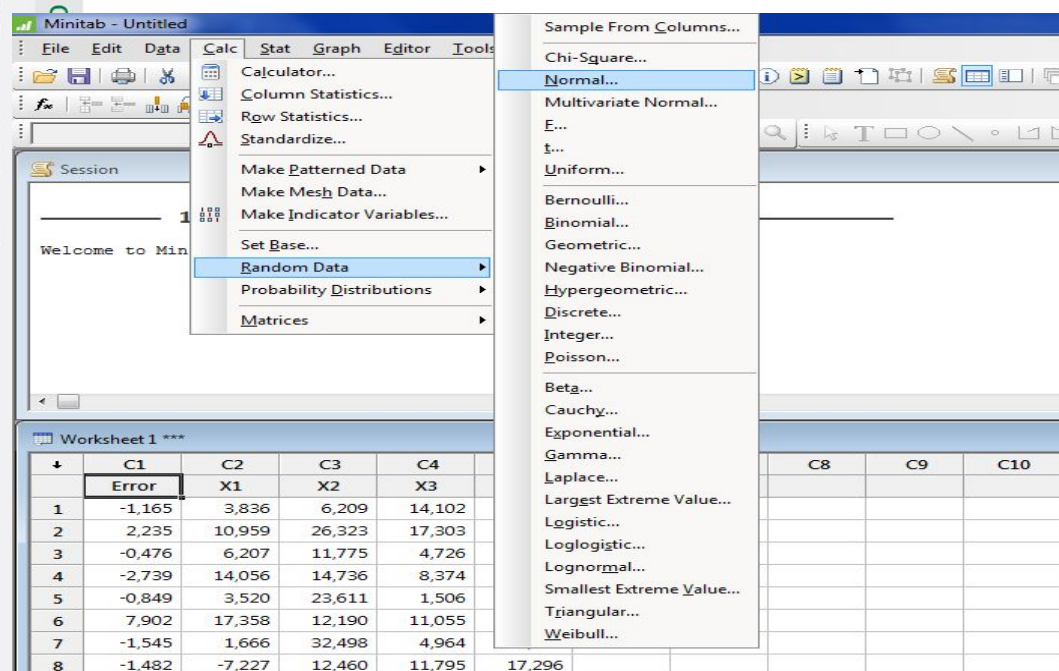
Gambar 3.1 Error yang Dibangkitkan

- Tentukan variabel bebas dan variabel terikat. Variabel bebas yang akan dibangkitkan sebanyak 3 variabel (X_1, X_2, X_3) dengan cara Klik Calc > Random data < Normal > pada kotak dialog “Number of rows of data to generate” tulis jumlah data yang akan dimasukkan (dalam penelitian ini data dibangkitkan sebanyak 63) dan pada kotak “store in column (s)” masukkan variabel.
- Tahap selanjutnya tentukan persamaan (Y) dengan cara klik Calc < Calculator < pada kolom “store in variabel” masukkan Y > kemudian buat persamaan regresi yang telah ditentukan. Dalam penelitian ini model yang digunakan adalah sebagai berikut:

$$\hat{Y} = 15,154 + 0,259 X_1 + 0,283 X_2 + 0,167 X_3 \quad (3.2)$$

- Tahap selanjutnya lakukan uji regresi berganda, dengan cara klik stat > regression > regression > fit regression models > pada kotak dialog masukkan variabel Y ke kotak “responses” dan masukkan semua variabel X ke kotak “continous predictor”.
- Kemudian klik Ok.
- Kemudian amati koefisien determinasi (R -Square / Adj R -Square dan variabel yang signifikan.

Adapun data simulasi yang digunakan selengkapnya disajikan pada Gambar 3.2 berikut:



Gambar 3.2 Prosedur untuk *Generated Data*

3.3 Pembentukan Model *Rough-Regresi*

Data yang telah diperoleh dari tahap simulasi selanjutnya dilakukan pemodelan dengan menggunakan *rough-regresi*. Langkah-langkah yang dilakukan adalah sebagai berikut :

- Melakukan uji regresi menggunakan data sebelum reduksi seperti yang dijelaskan pada Sub-Bab 2.1.1 hingga terbentuk model regresi.
- Pembentukan *decision system* yang berisikan atribut kondisi dan atribut keputusan. Tahap ini juga dilakukan untuk pembentukan kriteria berdasarkan kenormalan variabel, seperti yang telah disajikan pada Tabel 2.1 dan 2.2.
- Mengelompokkan data berdasarkan atribut kondisi dan atribut keputusan. Pada tahap ini dilakukan perbandingan antar data yang berbeda atribut kondisi dan atribut keputusan.
- Mengeliminasi atau mereduksi data.
- Melakukan uji regresi menggunakan data setelah direduksi.

3.4 Penggabungan *Outlier* dan *Unclassified* Data

Setelah diidentifikasi data yang merupakan *outlier* dengan melakukan uji regresi dan *unclassified* data menggunakan uji *rough sets* dengan memperhatikan nilai *R-Square* dari masing-masing percobaan yang dilakukan, maka langkah selanjutnya yaitu menggabungkan antara data *outlier* dan *unclassified* sehingga diperoleh data yang tidak bagus. Kemudian data yang tidak bagus tersebut akan digunakan untuk mengeliminasi data awal lalu dilakukan uji regresi dengan memperhatikan nilai *R-Square* untuk setiap percobaan.

3.5 Uji *Paired Samples T Test*

Pengujian ini dilakukan untuk melihat efek penggabungan *oulier* dan *unclassified* data. Adapun langkah-langkah pengujiannya sebagai berikut:

- Menyusun hipotesis nol dan alternatifnya, (H_0 dan H_1)
- Menentukan taraf signifikansi atau taraf nyata (α). Lawan taraf signifikansi adalah taraf kepercayaan $(1 - \alpha) = (1 - p)$.

Pengujian menggunakan uji dua sisi dengan tingkat signifikansi $\alpha = 5\%$.

- Menentukan t_{hitung}

Dari tabel *paired samples test* dapat diketahui nilai t_{hitung} .

- Menentukan t_{tabel}

Tabel distribusi t di cari pada $\alpha = 5\% : 2 = 2.5\%$ (uji 2 sisi) dengan derajat kebebasan $df = n - 1$. Dengan pengujian 2 sisi (signifikansi = 0.025) hasil t_{tabel} diperoleh dengan melihat tabel distribusi t .

- Kriteria Pengujian

H_0 ditolak jika $-t_{hitung} < -t_{tabel}$

Berdasarkan probabilitas:

H_0 ditolak jika $P_{value} < 0.05$

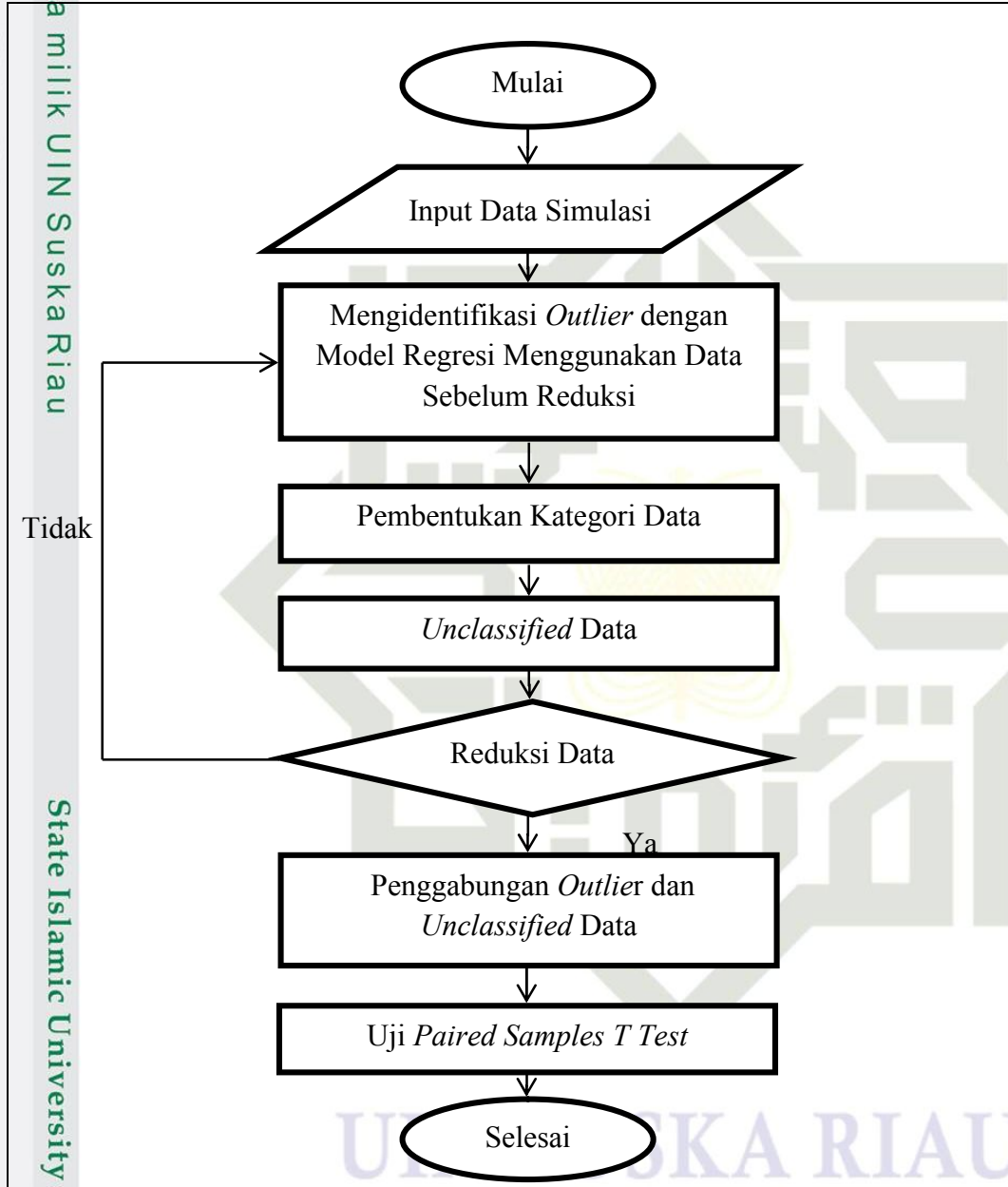
- Membandingkan t_{hitung} dengan t_{tabel} dan probabilitas
- Kesimpulan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.6 Tahap Evaluasi Model

Langkah-langkah persiapan data sampai pembentukan model diatas dapat digambarkan dalam *flowchart* sebagai berikut:



Gambar 3.3 *Flowchart* Implementasi Model *Rough-Regresi*

BAB V

PENUTUP

5.1 Kesimpulan

Pada penelitian ini, penggabungan antara *outlier* dan *unclassified* data pada pemodelan *rough*-regresi terdapat beberapa hasil dan manfaat yang bisa diambil diantaranya:

- a. Data simulasi yang dibangkitkan melalui *software* minitab 17.0 yaitu dengan menggunakan model regresi tingkat kemiskinan. Untuk melihat pengaruh penggabungan *outlier* dan *unclassified* maka dilakukan beberapa tahapan yaitu untuk mendeteksi adanya *outlier* dengan menggunakan hasil uji regresi linier berganda ataupun dapat dideteksi dengan *scatterplot/boxplot* diperoleh rata-rata *R-Square* dari seluruh percobaan sebesar 63.34% dan setelah dihilangkan data yang terindikasi sebagai *outlier*, rata-rata *R-Square* menjadi 64.28%. Selanjutnya, setelah dihilangkan data eliminasi (*unclassified* data) rata-rata *R-Square* berubah menjadi 69.34%. Dan dari hasil penggabungan *outlier* dan *unclassified* data rata-rata *R-Square* yang diperoleh mengalami kenaikan sebesar 71.78%. Selain itu, untuk melihat pengaruh penggabungan *outlier* dan *unclassified* data dapat digunakan uji *paired samples t test*. Dari hasil pengujian diperoleh kesimpulan bahwa terdapat pengaruh penggabungan *outlier* dan *unclassified* data terhadap *R-Square* masing-masing percobaan.
- b. Implementasi data real tingkat kemiskinan. Setelah hipotesis pada data simulasi berhasil, maka akan diterapkan ke dalam data real dengan tahapan yang sama seperti data simulasi. Sehingga diperoleh *R-Square* setelah *outlier* dihilangkan sebesar 48.41%, setelah *unclassified* data dihilangkan sebesar 84.70% dan penggabungan antara *outlier* dan *unclassified* data sebesar 87.69%.
- c. Berdasarkan hasil data simulasi dan data real tingkat kemiskinan, menunjukkan bahwa hipotesis pada data simulasi dapat bekerja dengan

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

baik pada data real. Hal ini berarti *R-Square* akan semakin meningkat jika dilakukan penggabungan *outlier* dan *unclassified* data.

Saran

Dalam penelitian ini telah dibahas mengenai data simulasi dan implementasinya ke data real dengan melihat *R-Square* dari masing-masing percobaan dan menentukan bahwa penggabungan antara *outlier* dan *unclassified* lebih baik dalam data simulasi maupun data real. Sehingga pembaca dapat menggunakan prosedur penggabungan data *outlier* dan *unclassified* dalam menghilangkan data yang membuat hasil yang diperoleh kurang baik.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR PUSTAKA

- Azwar, S. *"Penyusunan Skala Psikologi"*. Edisi Kedua. Yogyakarta: Pustaka Pelajar. 2012.
- Draper, N. dan Smith, H. *"Analisis Regresi Terapan"*. Edisi Kedua. Terjemahan Oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta. 1992.
- Efendi, R, Deris, M. M., Mu'at, S. "Decisive Criteria on Financial Literacy Using Rough-Regression Model for University Students". *Proceeding International Conference on Industrial, Enterprize, System Engineering (ICOIESE)*. IEEE Xplore. (Forthcoming). 2017
- Efendi, R., Deris, M. M. "Decision Making Model for Dominant Criteria Affecting Cholesterol Level Based On Rough-Regression". *Proceeding on Advances in Intelligent Systems and Computing*. Springer. (Forthcoming). Hal. 243-251. 2018.
- Efendi, R., Samsudin, N.A., Deris, M.M. "Medipre: Medical diagnosis prediction using rough-regression approximation". *ACM Proceeding on High Compilation, Computing and Communications*. Hal. 35-39. 2018a.
- Efendi, R., Mindiyarti, N., Hasanah, M., Suryani, I., Aryani, F., Fitriani, Y., Erlin. "Model Prediksi Fertilitas Menggunakan Metode Regresi, Rough Sets, dan Rough-Regresi". *Prosiding Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI-10)*. Hal. 666-672. 2018b.
- Gujarati, N.D. *"Basic Econometrics"*. 4th ed. McGraw-Hill Companies, New York. 2003.
- Gujarati, N.D. *"Dasar-Dasar Ekonometrika"*. Erlangga, Jakarta. 2006.
- Hartama, D dan Hartono. "Analisis Kinerja Dosen STMIK IBBI Dengan Menggunakan Metode Rough Set". *Jurnal Nasional Teknologi Informasi dan Multimedia*. Vol.4, No.1, hal. 49-54. 2016.
- Hasan, Iqbal. *"Pokok-pokok Materi Metodologi Penelitian dan Aplikasinya"*. Jakarta : Ghalia Indonesia. 2002.
- Herdiyatomoko, Hendrik Fery. "Reduksi Atribut Berdasar Matrik Discernibility Teori Rough Set Dalam Simulasi Bencana Kebakaran". *Jurnal SIMETRIS*. Vol. 8, No. 1, hal. 147-154, April 2017.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Herni, Ruri., & Abdurakhman. "Pengaruh Suatu Data Observasi Dalam Mengestimasi Parameter Model Regresi". *Jurnal Matematika dan Komputer*. Vol.5, No.3, hal. 161-167. 2002.
- Imawati. "Penerapan Teknik Artificial Intelligent Rough Set Untuk Mendukung Keputusan Pada Proses Pemeriksaan Kondisi Penjualan Barang Pada Toko Silungkang ART Center Padang", Vol.1 No 1 ISSN: 2356-0010. 2014.
- Jamaris, M. "Implementasi Metode Rough Set Untuk Menentukan Kelayakan Bantuan Dana Hibah Fasilitas Rumah Ibadah", Vol.2 N0.2 hal 161-170. 2017.
- Paludi, S. "Identifikasi dan Pengaruh Keberadaan Data Pencilan (Outlier)". *Majalah Ilmiah Panorama Nusantara*, edisi VI. 2009.
- Pawlak, Z. "Rough Set Theory and Its Applications". *Journal of Telecommunications and Information Technology*, 3, 7-10. 2002.
- Prajana, A. "Penerapan Teori Rough Set Untuk Memprediksi Tingkat Kelulusan Siswa Dalam Ujian Nasional Pada SMA Negeri 5 Kota Banda Aceh". *Journal of Islamic Science and Technology*. Vol.2, No.1. 2016.
- Rissino, S., Torres, L. "Rough set theory-fundamental concepts, principals, data extraction, and Knowledge Discovery in Real Life App". *Inform*, pp. 35-58. 2009.
- Santoso, S. "Buku Latihan SPSS Statistik Parametrik". Elex Media Komputindo, Jakarta. 2001.
- Seber, George A. F. & Lee, Alan J. "Linear Regression Analysis". New York: John Wiley & Sons. 2003.
- Sembiring, R. "Analisis Regresi edisi Kedua". Bandung: ITTB Bandung. 2003.
- Siahaan, P. "Penelitian Operasional : Teori dan Praktek". Jakarta : Universitas Indonesia Press. 1987.
- Soemarti. "Pencilan (Outlier)". Makalah Statistika FMIPA Universitas Padjajaran. Bandung. Tersedia: [http://resource.unpad.ac.id/unpad-content/apload/publikasi_dosen/Outlier\(Pencilan\).pdf](http://resource.unpad.ac.id/unpad-content/apload/publikasi_dosen/Outlier(Pencilan).pdf). 2007.
- Srinidi, I. G, Ayu Made. "Pengaruh Outlier Terhadap Estimator Parameter dan Metode Regresi Robust". *Prosiding Konferensi Nasional Matematika XVII*, Surabaya: 11-14 Juni . 2014.
- Sugiyono. "Metode Penelitian Bisnis", CV Alfabeta, Bandung. 1999.

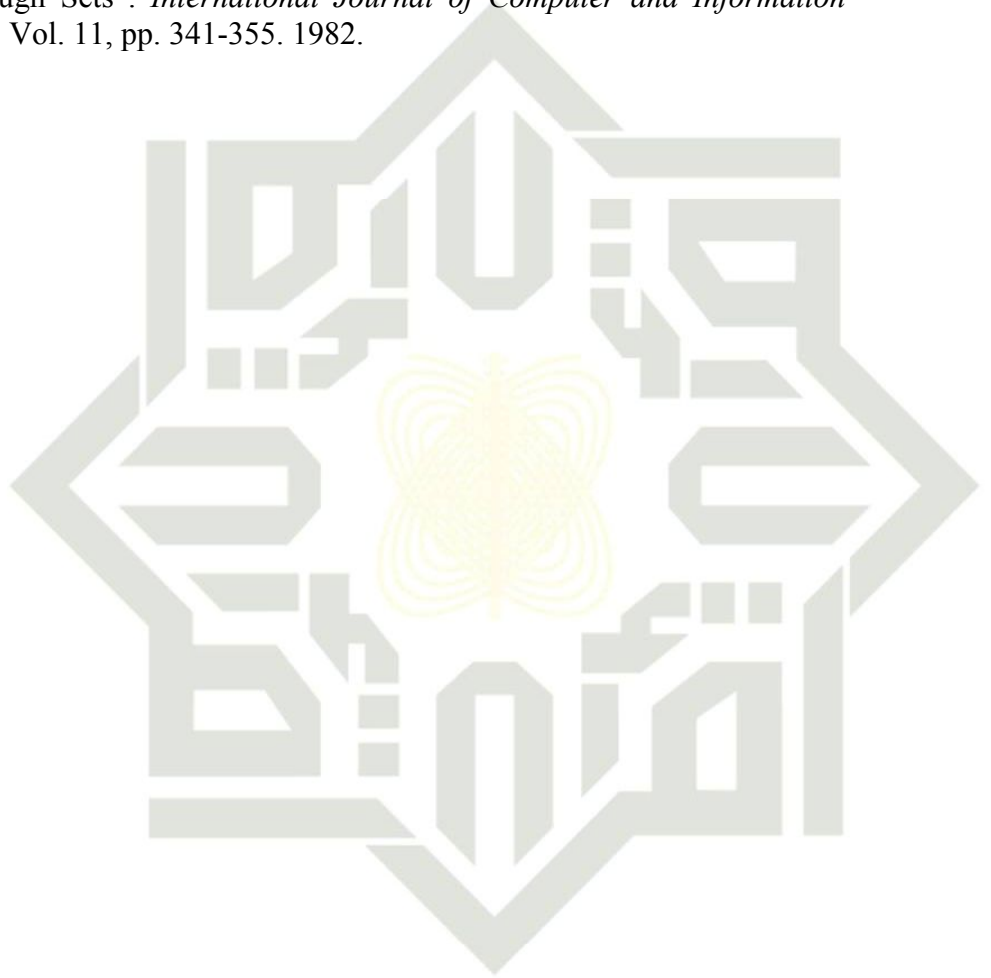
Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Syahputra, R., Muhammad. "Aplikasi Rough Set Untuk Memprediksi Prestasi Calon Anggota Kelompok Programming (Studi Kasus: STMIK Pelita Nusantara)". *Jurnal Mantik Penusa*. Vol.17, No.1, ISSN: 2088-3943. 2015.

Widarjono, A. "*Ekonomtrika Teori dan Aplikasi*". Yogyakarta: Ekonisia FE UII. 2007

Z. Pawlak. "Rough Sets". *International Journal of Computer and Information Sciences*. Vol. 11, pp. 341-355. 1982.



UIN SUSKA RIAU

LAMPIRAN A

Data Simulasi Tingkat Kemiskinan

No. Responden	Error	Inflasi (X_1)	UMR (X_2)	Pengangguran (X_3)	Kemiskinan (Y)
R-01	-1.165	3.836	6.209	14.102	19.095
R-02	2.235	10.959	26.323	17.303	30.566
R-03	-0.476	6.207	11.775	4.726	20.407
R-04	-2.739	14.056	14.736	8.374	21.624
R-05	-0.849	3.520	23.611	1.506	22.150
R-06	7.902	17.358	12.190	11.055	32.848
R-07	-1.545	1.666	32.498	4.964	24.067
R-08	-1.482	-7.227	12.460	11.795	17.296
R-09	-6.056	9.980	-13.104	11.593	9.911
R-10	-2.136	9.599	26.270	14.786	25.407
R-11	1.440	3.663	0.307	18.681	20.750
R-12	1.299	13.046	13.933	9.705	25.396
R-13	1.877	1.770	18.139	17.399	25.528
R-14	-4.907	-1.110	17.476	6.801	16.041
R-15	3.576	-1.938	15.418	17.348	25.488
R-16	1.890	29.477	12.488	18.447	31.293
R-17	-0.515	2.005	24.046	10.302	23.684
R-18	-1.649	14.726	14.403	13.271	23.612
R-19	5.029	16.174	20.504	16.402	32.914
R-20	4.002	-0.871	10.868	7.349	23.234
R-21	4.596	6.151	12.512	5.129	25.741
R-22	0.745	13.483	1.522	4.806	20.625
R-23	-1.303	3.418	13.367	15.560	21.118
R-24	-3.851	22.550	17.355	10.260	23.768
R-25	-0.191	7.554	3.700	10.107	19.655
R-26	5.772	7.071	1.322	12.025	25.139

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

R-27	-1.360	7.547	13.346	11.944	21.520
R-28	-3.078	12.524	9.800	12.289	20.146
R-29	-4.201	9.169	18.981	-0.264	18.655
R-30	3.579	6.013	4.944	13.856	24.003
R-31	-0.868	5.769	24.318	1.712	22.948
R-32	-2.204	7.523	4.342	7.159	17.323
R-33	2.663	12.184	8.670	7.154	24.621
R-34	-0.102	-6.998	14.048	15.474	19.799
R-35	-2.394	15.314	21.101	7.452	23.942
R-36	0.801	8.549	9.304	-0.128	20.780
R-37	-10.189	9.229	15.035	1.112	11.796
R-38	1.474	1.592	7.884	14.039	21.617
R-39	6.894	9.943	-12.662	6.955	22.201
R-40	1.979	6.472	7.883	11.946	23.035
R-41	6.404	-0.331	12.023	15.568	27.474
R-42	0.182	17.877	14.532	1.317	24.299
R-43	-4.679	0.650	17.376	5.330	16.451
R-44	2.101	2.575	19.280	11.453	25.291
R-45	0.028	8.057	2.155	4.796	18.679
R-46	7.169	14.530	14.288	13.068	32.312
R-47	-1.250	4.139	10.699	16.603	20.776
R-48	-0.230	0.846	5.168	9.176	18.138
R-49	-0.216	25.263	13.701	7.510	26.613
R-50	1.624	14.313	10.379	14.723	25.881
R-51	-4.711	3.729	13.343	8.940	16.677
R-52	-1.282	5.897	21.451	11.878	23.454
R-53	6.122	6.310	4.675	4.297	24.951
R-54	2.507	15.456	11.295	5.106	25.713
R-55	0.894	7.139	4.419	20.563	22.581
R-56	2.614	6.706	29.635	18.434	30.970

R-57	-6.560	11.701	10.739	11.056	16.510
R-58	-2.125	17.369	10.216	6.045	21.428
R-59	1.382	10.636	24.529	2.677	26.680
R-60	-3.135	2.144	11.950	11.585	17.891
R-61	-3.799	1.992	10.588	15.995	17.538
R-62	1.312	16.852	8.152	12.329	25.196
R-63	2.753	1.862	4.927	17.880	22.770

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LAMPIRAN B

Hasil Reduksi Data Simulasi Tingkat Kemiskinan

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Responden	X1	X2	X3	Y						
3	2	Sedang	Tinggi	Tinggi	Tinggi						
4	5	Sedang	Tinggi	Rendah	Sedang						
5	7	Sedang	Tinggi	Rendah	Sedang						
6	10	Sedang	Tinggi	Sedang	Sedang						
7	11	Sedang	Rendah	Tinggi	Sedang						
8	13	Sedang	Sedang	Tinggi	Sedang						
9	15	Rendah	Sedang	Tinggi	Sedang						
10	16	Tinggi	Sedang	Tinggi	Tinggi						
11	17	Sedang	Tinggi	Sedang	Sedang						
12	19	Tinggi	Sedang	Tinggi	Tinggi						
13	22	Sedang	Rendah	Rendah	Sedang						
14	31	Sedang	Tinggi	Rendah	Sedang						
15	35	Tinggi	Tinggi	Sedang	Sedang						
16	42	Tinggi	Sedang	Rendah	Sedang						
17	45	Sedang	Rendah	Rendah	Sedang						
18	47	Sedang	Sedang	Tinggi	Sedang						
19											
20											
21											
22											
23											
24											
25											

Hasil Eliminasi Data Simulasi Tingkat Kemiskinan

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Responden	X1	X2	X3	Y						
3	1	Sedang	Sedang	Sedang	Sedang						
4	3	Sedang	Sedang	Rendah	Sedang						
5	4	Sedang	Sedang	Sedang	Sedang						
6	6	Tinggi	Sedang	Sedang	Tinggi						
7	8	Rendah	Sedang	Sedang	Rendah						
8	9	Sedang	Rendah	Sedang	Rendah						
9	12	Sedang	Sedang	Sedang	Sedang						
10	14	Rendah	Sedang	Sedang	Rendah						
11	18	Sedang	Sedang	Sedang	Sedang						
12	20	Rendah	Sedang	Sedang	Sedang						
13	21	Sedang	Sedang	Sedang	Sedang						
14	23	Sedang	Sedang	Sedang	Sedang						
15	24	Tinggi	Sedang	Sedang	Sedang						
16	25	Sedang	Rendah	Sedang	Sedang						
17	26	Sedang	Rendah	Sedang	Sedang						
18	27	Sedang	Sedang	Sedang	Sedang						
19	28	Sedang	Sedang	Sedang	Sedang						
20	29	Sedang	Sedang	Rendah	Sedang						
21	30	Sedang	Sedang	Sedang	Sedang						
22	32	Sedang	Sedang	Sedang	Rendah						
23	33	Sedang	Sedang	Sedang	Sedang						
24	34	Rendah	Sedang	Sedang	Sedang						
25	36	Sedang	Sedang	Rendah	Sedang						

Hak Cipta Dilindungi Undang-Undang

R-Square dari Outlier, Unclassified Data serta Penggabungan Outlier dan Unclassified Data

No. Percobaan	Outlier (O)	Unclassified Data (U)	(O + U)
P-01	47.79%	81.87%	77.65%
P-02	62.65%	60.46%	60.46%
P-03	66.99%	75.66%	75.66%
P-04	60.87%	74.32%	63.19%
P-05	68.35%	54.62%	71.39%
P-06	65.73%	94.13%	94.13%
P-07	57.99%	46.12%	91.37%
P-08	74.59%	51.32%	51.32%
P-09	76.88%	81.90%	88.91%
P-10	78.38%	93.37%	93.37%
P-11	61.96%	59.51%	60.82%
P-12	60.95%	73.54%	87.47%
P-13	77.11%	51.80%	67.64%
P-14	69.79%	86.87%	60.26%
P-15	51.60%	90.70%	93.49%
P-16	51.44%	85.24%	51.42%
P-17	70.88%	88.80%	62.42%
P-18	72.03%	87.37%	54.11%
P-19	70.14%	53.78%	59.56%
P-20	48.01%	49.92%	83.76%
P-21	60.03%	66.03%	85.32%
P-22	59.26%	93.99%	54.99%
P-23	60.18%	49.23%	80.84%
P-24	67.31%	74.24%	55.47%
P-25	72.18%	93.36%	64.07%
P-26	64.28%	47.68%	70.16%
P-27	67.07%	75.06%	86.62%

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

P-28	66.19%	73.51%	80.83%
P-29	52.50%	78.03%	61.20%
P-30	49.57%	46.54%	82.47%
P-31	69.46%	63.21%	54.69%
P-32	52.38%	49.30%	64.59%
P-33	74.67%	61.41%	72.74%
P-34	61.84%	85.31%	62.85%
P-35	56.72%	51.18%	66.20%
P-36	54.69%	61.15%	54.24%
P-37	72.68%	75.58%	76.41%
P-38	52.00%	71.51%	74.58%
P-39	66.21%	53.64%	68.38%
P-40	69.31%	80.08%	88.29%
P-41	74.95%	84.24%	86.70%
P-42	68.50%	84.71%	59.78%
P-43	57.76%	74.41%	54.39%
P-44	75.71%	61.38%	62.92%
P-45	56.53%	58.25%	72.48%
P-46	77.83%	68.99%	74.53%
P-47	65.13%	81.44%	60.69%
P-48	59.89%	53.22%	84.88%
P-49	67.97%	62.77%	90.70%
P-50	67.25%	46.31%	88.80%

LAMPIRAN C

Titik Persentase Distribusi t (df = 41 – 80)

Pr df	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010	0.001 0.002
41	0.68052	1.30254	1.68288	2.01954	2.42080	2.70118	3.30127
42	0.68038	1.30204	1.68195	2.01808	2.41847	2.69807	3.29595
43	0.68024	1.30155	1.68107	2.01669	2.41625	2.69510	3.29089
44	0.68011	1.30109	1.68023	2.01537	2.41413	2.69228	3.28607
45	0.67998	1.30085	1.67943	2.01410	2.41212	2.68959	3.28148
46	0.67986	1.30023	1.67866	2.01290	2.41019	2.68701	3.27710
47	0.67975	1.29982	1.67793	2.01174	2.40835	2.68456	3.27291
48	0.67964	1.29944	1.67722	2.01063	2.40658	2.68220	3.26891
49	0.67953	1.29907	1.67655	2.00958	2.40489	2.67995	3.26508
50	0.67943	1.29871	1.67591	2.00856	2.40327	2.67779	3.26141
51	0.67933	1.29837	1.67528	2.00758	2.40172	2.67572	3.25789
52	0.67924	1.29805	1.67469	2.00665	2.40022	2.67373	3.25451
53	0.67915	1.29773	1.67412	2.00575	2.39879	2.67182	3.25127
54	0.67906	1.29743	1.67356	2.00488	2.39741	2.66998	3.24815
55	0.67898	1.29713	1.67303	2.00404	2.39608	2.66822	3.24515
56	0.67890	1.29685	1.67252	2.00324	2.39480	2.66651	3.24226
57	0.67882	1.29658	1.67203	2.00247	2.39357	2.66487	3.23948
58	0.67874	1.29632	1.67155	2.00172	2.39238	2.66329	3.23680
59	0.67867	1.29607	1.67109	2.00100	2.39123	2.66176	3.23421
60	0.67860	1.29582	1.67065	2.00030	2.39012	2.66028	3.23171
61	0.67853	1.29558	1.67022	1.99962	2.38905	2.65886	3.22930
62	0.67847	1.29536	1.66980	1.99897	2.38801	2.65748	3.22696
63	0.67840	1.29513	1.66940	1.99834	2.38701	2.65615	3.22471
64	0.67834	1.29492	1.66901	1.99773	2.38604	2.65485	3.22253
65	0.67828	1.29471	1.66864	1.99714	2.38510	2.65360	3.22041
66	0.67823	1.29451	1.66827	1.99656	2.38419	2.65239	3.21837
67	0.67817	1.29432	1.66792	1.99601	2.38330	2.65122	3.21639
68	0.67811	1.29413	1.66757	1.99547	2.38245	2.65008	3.21446
69	0.67806	1.29394	1.66724	1.99495	2.38161	2.64898	3.21260
70	0.67801	1.29376	1.66691	1.99444	2.38081	2.64790	3.21079
71	0.67796	1.29359	1.66660	1.99394	2.38002	2.64686	3.20903
72	0.67791	1.29342	1.66629	1.99346	2.37926	2.64585	3.20733
73	0.67787	1.29326	1.66600	1.99300	2.37852	2.64487	3.20567
74	0.67782	1.29310	1.66571	1.99254	2.37780	2.64391	3.20406
75	0.67778	1.29294	1.66543	1.99210	2.37710	2.64298	3.20249

- Hak Cipta Dilindungi Undang-Undang**
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



DAFTAR RIWAYAT HIDUP

Penulis dilahirkan di Kota Pekanbaru tanggal 05 Mei 1998, sebagai anak kedua dari empat bersaudara pasangan Bapak Syamsul Bahri dan Ibu Nurbayati. Penulis menyelesaikan Pendidikan Formal Sekolah Dasar di SD Negeri 031 Tampan tahun 2010. Pada tahun 2013 penulis menyelesaikan Pendidikan Lanjutan Tingkat Pertama di SMP Negeri 20 Pekanbaru dan menyelesaikan Pendidikan Menengah Atas dengan jurusan Ilmu Pengetahuan Alam (IPA) di SMA Negeri 12 pada tahun 2016 di Kota Pekanbaru.

Tahun 2016 penulis melanjutkan Pendidikan ke Perguruan Tinggi di Universitas Islam Negeri Sultan Syarif Kasim Riau dan lulus di Fakultas Sains dan Teknologi dengan Jurusan Matematika. Pada bulan Februari 2019, penulis melaksanakan Kerja Praktek di UPT. Laboratorium Bahan Konstruksi Dinas Pekerjaan Umum dan Penataan Ruang Provinsi Riau, dengan judul **“Pengaruh Kualitas Pelayanan dan Kinerja Karyawan Terhadap Kepuasan Pelanggan UPT Laboratorium Bahan Konstruksi Menggunakan Model Regresi Linier Berganda”** yang dibimbing oleh Ibu Sri Basriati, M.Sc. yang diseminarkan pada tanggal 03 Juli 2019. Bulan Juli-Agustus 2019 penulis mengikuti Kuliah Kerja Nyata (KKN) di Kabupaten Kampar, Kecamatan Gunung Sahilan, Desa Gunung Mulya. Penulis dinyatakan lulus ujian sarjana pada tanggal Desember 2019 dengan judul Tugas Akhir **“Efek Penggabungan *Outlier* dan *Unclassified Data* Pada Pemodelan *Rough-Regresi* (Studi Kasus: Tiga Variabel Bebas Numerik”** dengan dosen pembimbing Bapak Dr. Riswan Efendi, M.Sc.